Asset Pricing Theory and Tests

Robert R. Grauer*
Endowed University Professor
Faculty of Business Administration
Simon Fraser University
8888 University Drive
Burnaby, British Columbia, Canada V5A 1S6
Telephone: (604) 291-3722   FAX: (604) 291-4920
E-mail: grauer@sfu.ca

October 2002

Acknowledgements

Foreword  Richard Roll

Introduction  Robert R. Grauer

PART I  AN OVERVIEW OF ASSET PRICING THEORY  [169 pp]


PART II  PRE-1990 TESTS OF THE MEAN-VARIANCE CAPITAL ASSET PRICING MODEL  [151 pp]


*Name Index* [433 pp]
Contents

Acknowledgements

An Introduction by the editor to both volumes appears in Volume I


PART V TESTS OF THE LINEAR RISK TOLERANCE CAPMS  [23 pp]

PART VI TESTS OF THE CONSUMPTION-BASED CAPM  [85 pp]

PART VII TESTS OF THE ARBITRAGE PRICING THEORY  [66 pp]

PART VIII TESTS OF AN INVESTMENT-BASED CAPM USING THE GENERALIZED METHOD OF MOMENTS  [78 pp]

*Name Index*  [580 pp]
Acknowledgements

Anyone who assembles a book of readings quickly realizes that he will either miss or simply have to leave out major papers and even whole areas of the literature. I have not included any readings on skewness or stochastic dominance models or on mean-variance spanning, for example. I can only apologize to those who may feel slighted.

For the most part, I have endeavored to focus on basic ideas as opposed to more technical material. I have also attempted to capture the invigorating controversy that surrounds both the design of empirical tests and the interpretation of the results. Much of the controversy is in the mainstream. Whether the capital asset pricing model has been erroneously rejected because of data snooping, survivorship bias, or poor choice of market proxies; whether asset pricing is rational and conforms to a three-factor intertemporal capital asset pricing model or arbitrage pricing theory that does not reduce to the traditional capital asset pricing model; or whether investor irrationality prevents the three-factor model from collapsing to the capital asset pricing model have been vigorously debated in the recent literature. But more basic issues have been all but ignored. Do the tests of risk-return tradeoffs mean anything if we cannot observe the true market portfolio? Do the most basic assumptions of homogeneous beliefs, of \textit{ex post} returns being drawn from \textit{ex ante} return distributions, of testing a single-period model where returns are assumed to be stationary make sense? I don't think they should be. Finally, although I have focused on what I think are some of the most important articles in the area, I also refer to a number of working papers that illustrate how the controversy surrounding the empirical results continues unabated. Only the passage of time will determine whether any of the working papers will become classics.

I gratefully acknowledge financial support from the Social Sciences and Research Council of Canada and the most capable assistance of Poh Chung Fong, Christopher Fong, and Jan MacLellan. Without implicating them, I also thank Stephane Chretien, John Cochrane, Raymond Kan, Peter Kennedy, Mark Rubinstein, William Sharpe, and especially Peter Klein for their most helpful comments. I am most indebted to Richard Roll, however, whose insightful comments and suggestions greatly improved the manuscript.
Introduction

Asset pricing lies at the heart of financial economics. It is difficult to imagine how we would teach corporate finance or investments without a theory of value. Likewise, it is hard to conceive how we could make the efficient markets hypothesis operational without some underlying model of how assets are priced. Without an asset pricing model, everyday applications like risk analysis, cost-of-capital calculations, or measuring the performance of mutual funds would be an adventure. In this volume, I compile readings on the theory of asset pricing as well as on the tests of that theory. I focus primarily on models that are based on economic rationality in an uncertain environment. A separate volume in this series, edited by Hersh Shefrin, is devoted to behavioral finance, and three volumes, edited by Stephen Ross, are devoted to the debt market.

Most of the theory on asset pricing has been in place for well over twenty years. The 1960's saw the birth of the single-period mean-variance (MV) capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965), Mossin (1966), and Black (1972). The 1970's brought the added richness of Merton's (1973) intertemporal capital asset pricing model (ICAPM), Rubinstein's (1974) single-period linear risk tolerance (LRT) models, Ross' (1976, 1977) Arbitrage Pricing Theory (APT), and the intertemporal consumption-based model (CCAPM) of Rubinstein (1976), Lucas (1978), and Breeden (1979). On the other hand, while the first extensive tests of the MV CAPM appeared in the early 1970's, debates over which of the models holds empirically have raged unabated since Roll's (1977) critique. Given these often-heated, on-going, and intensely interesting debates on what (if anything) the empirical results mean, I focus primarily on the testing literature.

On the theory side, I focus on the models that have been subject to empirical testing. I do not, for example, trace the existence of a stochastic discount factor to the law of one price or the existence of a positive stochastic discount factor to the absence of arbitrage. Moreover, I limit the scope to absolute pricing, where an asset is priced by its exposure to fundamental sources of risk, rather than relative pricing, where an asset is priced relative to given values of some other assets.

On the empirical side, I focus on the basic ideas rather than the details of different statistical methods as, when all is said and done, they really attempt to do the same thing. Cochrane (2001a, p. xvi) sums up this point of view nicely:

A wide variety of methods are popular, including time-series and cross-sectional regressions, and methods based on generalized methods of moments (GMM) and maximum likelihood. However, in the end all of these apparently different approaches do the same thing: they pick free parameters of the model to make it fit best, which usually means to minimize pricing errors; and they evaluate the models by examining how big those pricing errors are.

In the language of expected returns, this translates into testing whether average returns are equal to the expected returns predicted by the model under consideration. Instead, I emphasize critiques of the tests that arise from logical rather than statistical considerations, as I believe they raise the most fundamental doubts about the empirical evidence.

Where possible I have chosen articles to be included in this volume that provide the reader with perspective and intuition. For example, I chose Sharpe (1991) over Sharpe (1964) to
introduce the MV CAPM. For the APT, I chose Ross (1977) over Ross (1976). Finally, I include excellent survey articles by Jensen (1972), Fama (1991), and Campbell (2000).

The paper proceeds as follows. Section I provides a brief overview of asset pricing theory: the mean-variance CAPM, the linear risk tolerance CAPMs, the intertemporal CAPM, the consumption-based CAPM, and the arbitrage pricing theory. It portrays the models as either tastes-based or beliefs-based. The common characteristic of the beliefs-based models is that the pricing equations follow directly from the optimality conditions of portfolio choice. Both the tastes-based and beliefs-based pricing equations are also related to the stochastic discount factor methodology.

Sections II through VIII are concerned with empirical testing. Sections II-IV focus on the mean-variance model. Section II discusses the pre-1990 literature. It begins with an examination of the testable implications of the model together with an assessment of some of the fundamental, but seldom discussed, assumptions employed in developing the tests. For example, why do we not test for two-fund separation? Why don't we test whether the minimum-variance frontier contains portfolios with all positive weights? Does the assumption of homogeneous beliefs make sense? The central time-series and cross-sectional tests that focus on the security market line relationship and the multivariate tests of the efficiency of the 'market' portfolio are then summarized. The rest of the section recaps the published research and the fundamental critiques of it. Do the tests of risk-return tradeoffs mean anything if we cannot observe the true market portfolio? Do the most basic assumptions of \textit{ex post} returns being drawn from \textit{ex ante} return distributions or of testing a single-period model where returns are assumed to be stationary make sense? Sections III and IV review the post-1990 literature. Section III discusses CAPM anomalies and Fama and French's three-factor model. Section IV focuses on the exciting controversy that surrounds both the design of empirical tests and the interpretation of the results. Questions of whether the capital asset pricing model has been erroneously rejected because of data snooping, survivorship bias, or poor choice of market proxies; whether asset pricing is rational and conforms to a three-factor intertemporal capital asset pricing model or arbitrage pricing theory that does not reduce to the traditional capital asset pricing model; or whether investor irrationality prevents the three-factor model from collapsing to the capital asset pricing model form the basis of a gripping on-going debate. The section continues with discussions of possible roles for liquidity and taxes and conditional models. It concludes with an examination of the robustness of the empirical results. Section V employs a numerical example to emphasize the difficulties associated with distinguishing between the power utility-based linear risk tolerance and mean-variance CAPMs. Sections VI-VIII discuss tests of the consumption-based CAPM, the Arbitrage Pricing Theory, and an investment-based CAPM using the generalized method of moments. Section IX contains a summary and concluding comments.

I An Overview of Asset Pricing Theory

Each of the asset-pricing models has its appeal. The single-period MV model is by far the best known. From a theoretical point of view, it represents an almost perfect blend of elegance and simplicity. Beta is an intuitively appealing measure of risk whether one argues it is an asset's contribution to total societal risk, i.e., the variance of the market portfolio, or the part of risk that cannot be diversified away. From an empirical point of view, the model appears to be readily testable. Betas are easily estimated from standard time-series regressions. And, a linear risk-return tradeoff seems to be tailor-made for empirical testing. Arguably, the expected utility-
based LRT models have a deeper theoretical appeal. However, they do not fit as nicely into a standard return generating model regression framework.

The theoretical appeal of Merton's ICAPM is that it allows investors to hedge against changes in the opportunity set, a motive which is clearly not present in single-period models. Moreover, from an empirical point of view, the ICAPM and the APT yield multifactor-pricing equations that allow for more than just market risk. As Fama (1991, page 1594) points out: 'The multifactor models are an empiricist's dream. They are off-the-shelf theories that can accommodate tests for cross-sectional relations between expected returns and the loadings of security returns on any set of factors that are correlated with returns.' However, Fama also notes (page 1595): 'There is an important caveat. The flexibility of the Chen, Roll, and Ross approach can be a trap. Since multifactor models offer at best vague predictions about the variables that are important in returns and expected returns, there is the danger that measured relations between returns and economic factors are spurious, the result of special features of a particular sample (factor dredging).'

Even so, the CCAPM is the most elegant of all the models. For example, Breeden's model summarizes all the incentives to hedge shifts in consumption and portfolio opportunities with a one-factor relation between expected returns and consumption betas. Unfortunately, the empirical evidence does not offer strong support for the model.

It is difficult to develop propositions about asset pricing that are both interesting and general without placing restrictions on tastes (investors' utility functions) or beliefs (investors' subjective probability estimates). Beliefs-based models include the single-period MV model, derived by assuming joint normal returns, Merton's continuous-time lognormality model, Breeden's continuous-time consumption-based model, and Ross' APT. Tastes-based models include the MV model, derived by assuming quadratic utility, as well as Rubinstein's single-period LRT models.

It is somewhat disconcerting, however, that the tastes-beliefs breakdown is not completely clean. There are three reasons for this. First, the same model can be developed in different ways. For example, the MV model can be derived in at least three different ways. As noted, the model can be based on the assumption of joint normality. Or, the riskfree asset version of the MV CAPM can be obtained as a special case of the LRT models where all investors have quadratic utility. Or, Roll and Ross (1980) argue that the real intuition for the MV CAPM is based on the ideas of diversifiable risk and nondiversifiable risk, which stem from Sharpe's (1963) diagonal model and his use of it in his original development of the CAPM. Second, while much of the early theoretical appeal of the APT is based on the fact it does not place restrictions on tastes (except that more is preferred to less), it causes problems for empiricists as the pricing equation holds only as an approximation. Yet, when Connor (1984) develops an exact APT pricing relation, his theory relies on the principles of competitive equilibrium rather than on arbitrage. Third, the waters are muddied further as empiricists are forced to make assumptions that are not required by theory. For example, normality is often assumed for statistical reasons. Moreover, the pricing equations are stated in terms of the ex ante expectations of investors while researchers can only observe ex post returns. Thus, many of the tests rely on a generating (factor) model to make the connection between ex post and ex ante returns.

While there are many ways to develop the models, I focus on derivations based on the optimality conditions of portfolio choice. I begin with the single-period models: the MV model,
the LRT models, the consumption-based model, and the stochastic discount factor model. I then proceed to the intertemporal investment and consumption models. The unifying idea is that the pricing equations are simply the first-order conditions of portfolio choice. The APT is discussed last, as its genesis from a beliefs-based factor model does not easily fit into the portfolio-choice framework.

Following Markowitz (1959), Sharpe (1970, 1991), and Best and Grauer (1990) the mean-variance problem is

$$\max [T_m(\mu' x + R_f x_f) - \frac{1}{2} x' \Sigma x] + \lambda (1 - t' x - x_f),$$  \hspace{1cm} (1)

where $T_m$ is the risk tolerance parameter of the representative investor, $\mu$ and $x$ are $n$-vectors containing the unity plus the expected rates of return on the $n$ assets and the portfolio weights, respectively, $R_f$ is unity plus the riskfree rate of interest, $x_f$ is the weight invested in the riskfree asset, $\Sigma$ is an $n \times n$ positive-definite covariance matrix of asset returns, $t$ is an $n$-vector containing ones, and $\lambda$ is the Lagrange multiplier associated with the (normalized) budget constraint. (For notational purposes, a prime indicates transposition. For example, $x'$ is the row vector corresponding to the column vector $x$.) The optimality conditions are

$$T_m \mu - \Sigma x - \lambda t = 0, \quad T_m R_f = \lambda, \quad t' x + x_f = 1. \quad \hspace{1cm} (2)$$

One can easily solve equation (2) for the representative investor's optimal portfolio, which in this case must be the market portfolio $x_m$. However, we simply note that rewriting the first part of the first-order conditions yields the security market line (SML)

$$\mu = \frac{\lambda}{T_m} t + \frac{1}{T_m} \Sigma x_m. \hspace{1cm} (3)$$

Let the $j$-th element of $\mu$ be $E(R_j)$ — unity plus the expected rate of return on asset $j$—and note that the $j$-th element of the vector $\Sigma x_m$ is equal to $\text{cov}(R_j, R_m)$—the covariance of the return on asset $j$ with the return on the market portfolio. Then equation (3) may be written in scalar notation either as

$$E(R_j) = R_f + \frac{E(R_m) - R_f}{\sigma_m^2} \text{cov}(R_j, R_m), \hspace{1cm} (4a)$$

or in the more familiar form as

$$E(R_j) = R_f + (E(R_m) - R_f) \beta_j, \hspace{1cm} (4b)$$

where $\sigma_m^2$ is the variance of the return on the market, $\beta_j = \text{cov}(R_j, R_m)/\sigma_m^2$, and the reciprocal of the risk tolerance parameter $T_m$ is equal to $(E(R_m) - R_f)/\sigma_m^2$.

Sharpe's Nobel Lecture (1991, Volume I, Chapter 1) develops the MV CAPM in a succinct manner that emphasizes the economic content of the theory as well as the model's relation to theories that extend, adapt, or compete with it. Sharpe then relaxes the assumption that investors can freely short sell risky assets. The considerably more complex results suggest a reduction in the efficiency with which risk can be allocated in an economy. Finally, in an example of practice
imitating theory, he discusses how stock market futures contracts help bring actual capital markets closer to those of the simple equilibrium theory, thereby significantly improving the efficiency of societal risk sharing.

The single-period expected utility-based portfolio selection problem, subject to the constraint that the sum of the dollar amounts invested in the risky assets and a riskfree asset are equal to initial wealth, is

$$E(u(w_1)) + \lambda (w_0 - \sum z_j + z_f),$$

(5)

where $E(.)$ denotes the expectations operator, $u(w_1)$ is the utility of wealth at time 1, $w_0$ is initial wealth, $z_j$ is the dollar amount invested in risky asset $j$, and $z_f$ is the amount borrowed or lent and $w_1 = \sum z_j R_j + z_f R_f$. The first-order conditions include

$$\lambda = \frac{1}{\lambda} \lambda E(u'(w_1) R_j), \text{ for all } j, \quad \text{ and } \quad w_0 = \sum z_j + z_f.$$

Rubinstein (1974, Volume I, Chapter 2), among others, using the fact that $E(\sigma Y) = E(X)E(Y) + \text{cov}(X,Y)$, shows that the optimality conditions may be written either as

$$E(u'(w_1) R_j) = \lambda, \text{ for all } j, \quad E(u'(w_1) R_f) = \lambda, \text{ and } \quad w_0 = \sum z_j + z_f.$$

or as

$$E(R_j) = R_j + \frac{1}{E(u'(w_1))} \text{cov}(R_j, -u'(w_1)).$$

He then shows that for the family of linear risk tolerance utility functions, where an aggregate investor can be constructed, the marginal utility of wealth can be stated in terms of potential observable data.

More specifically, Rubinstein develops alternative sets of sufficient conditions under which equilibrium security rates of return are determined as if there exists only identical individuals whose resources, beliefs, and tastes are a composite of the actual individuals in the economy. These expected utility-based valuation equations bridge the gap between the single-period MV model and multiperiod asset pricing models. The aggregation theorem relies on the separation property of the linear risk tolerance utility functions and the observation that under popular homogeneity assumptions regarding beliefs and tastes, even though the securities market may be incomplete, equilibrium rates of return are determined as if there were a complete (Arrow-Debreu) market. In addition to providing the valuation equations that are of primary interest to readers of this volume, Rubinstein uses the "as if complete markets" argument to show that in these economies both market exchange arrangements and production decisions are Pareto-optimal.

Instead of defining utility in terms of wealth, we can define it over current and future consumption. Following Cochrane (2001a), we consider the simplest problem. Assume the investor can freely buy or sell as much of a payoff $y_{t+1}$ as he wishes at a price $p_t$. Let $e$ be the value of his endowed consumption and denote by $\varepsilon$ the amount of the asset he chooses to buy.
Then the investor’s expected utility problem is
\[
u(c_t, c_{t+1}) = u(c_t) + \delta E(u(c_{t+1})� (7)
\]
subject to
\[
c_t = e_t - p_e \\
c_{t+1} = e_{t+1} + y_{t+1}e
\]
where \( \delta \) is the investor’s subjective discount factor that captures his tradeoff between current consumption and the expected value of future consumption. Substituting the constraints into the objective function and setting the derivative with respect to \( \varepsilon \) equal to zero, we obtain the first-order condition for an optimal consumption and portfolio choice
\[
p_t = E(\delta \frac{u'(c_{t+1})}{u'(c_t)} y_{t+1}). (8)
\]

If we define the stochastic discount factor (also known as the marginal rate of substitution, the pricing kernel, or the state-price density) to be
\[
m_{t+1} = \delta \frac{u'(c_{t+1})}{u'(c_t)}, (9)
\]
then the basic pricing formula, equation (8), can be expressed as
\[
p_t = E(m_{t+1}y_{t+1}). (10)
\]
If we further define \( R_{t+1} = y_{t+1} / p_t \), equation (10) can be rewritten in terms of returns
\[
E(m_{t+1}R_{t+1}) = 1. (11)
\]
More generally, equation (11) holds for any asset \( j \). Employing again the result that \( E(XY) = E(X)E(Y) + \text{cov}(X,Y) \), the result that the riskfree return \( R_f = 1 / E(m_{t+1}) \), and dropping the \( t+1 \) subscript from the \( j \)th asset’s return yields the basic pricing equation
\[
E(R_j) = \frac{1}{E(m)} \frac{1}{E(m)} \text{cov}(R_j, -m), (12a)
\]
or
\[
E(R_j) = R_f + \frac{1}{E(u'(c_{t+1}))} \text{cov}(R_j, -u'(c_{t+1})). (12b)
\]

Finally, let \( \rho = \text{cov}(m, R_j) / \sigma(m) \sigma(R_j) \) be the correlation between \( m \) and \( R_j \), where the standard deviation of the stochastic discount factor is written as \( \sigma(m) \) to distinguish it from the standard deviation of the return on the market, which is denoted as \( \sigma_m \). Then equation (12a) can be rewritten as
\[
E(R_j) = \frac{1}{E(m)} - \rho \frac{1}{E(m)} \sigma(m) \sigma(R_j). \]
Given that correlation coefficients are less than or equal to 1, this means that
\[
\left| \frac{E(R_j) - R_j}{\sigma(R_j)} \right| \leq \frac{\sigma(m)}{E(m)},
\] (13)
which is known as the Hansen-Jagannathan (1991) bound. With the assumption of either lognormality or continuous-time decision-making
\[
\left| \frac{E(R_j) - R_j}{\sigma(R_j)} \right| \leq \frac{\sigma(m)}{E(m)} \approx (1 - \gamma)\sigma(\Delta),
\] (14)
where \((1 - \gamma)\) is the Pratt-Arrow measure of relative risk aversion for a power utility investor given in equation (15) below, and \(\sigma(\Delta) = \ln c_{t+1} - \ln c_t\). Shortly, we will see how these results help to explain the equity premium puzzle that bedevils the CCAPM.

While the pricing equations may appear different, they have fundamental similarities. First, all the pricing equations follow from the optimality conditions of portfolio choice. Second, the riskfree rate and the risk premium are functions of investors' tastes (or attitudes toward risk). Third, the risk measures are determined by the covariance of an asset's returns with the representative investor's marginal utility. To be more specific, if we compare equations (6) and (12) we see that the single-period LRT models are just special cases of equation (12) with the stochastic discount factor \(m\) set equal to \(\lambda = \frac{w_u}{\lambda}(101)\). In case of quadratic utility, the aggregate investor's marginal utility of wealth is equal to \(\sigma(\Delta) = \ln c_{t+1} - \ln c_t\). In terms of returns, securities whose payoffs are positively correlated with wealth (consumption) have to promise higher expected returns to induce investors to hold them.

Merton (1973, Volume I, Chapter 3) extends the single-period CAPM to a multiperiod framework. He develops the model in continuous time where he assumes returns follow a lognormal diffusion process. Unlike the single-period model, expected returns are affected by the possibility of uncertain changes in future investment opportunities. In a special case, where one asset's return is perfectly negatively correlated with changes in the riskfree rate of interest, he develops a three-fund CAPM. In this case, expected returns are a function of market risk and the risk of bearing unfavorable shifts in the investment opportunity set. More generally, Merton's insights allow empiricists to search for factors or state variables that may signal changes in investment opportunities. Fama and French's (1993) distress factor and Ferson and Harvey's (1999) use of dividend yields, term-structure variables, and riskfree interest rates as conditioning variables that are intended to capture changes in the opportunity set are prime examples of this type of research.

Merton's ICAPM and the Black and Scholes' (1973) option pricing model are rightly viewed as among the most important contributions to financial economics. Both rely on continuous-time decision-making coupled with the assumption of Brownian motion. Rubinstein (1976, Volume I,
Chapter 4) develops a discrete-time CCAPM and an alternative way of pricing options. The paper contains more than one model. The first assumes joint normality of the cash flow of a security with aggregate consumption at each date. No assumptions are made about utility, other than the usual assumptions of time additivity and the signs of the first and second derivatives. Using the Stein-Rubinstein lemma, Rubinstein derives a discrete-time CCAPM. However, to apply the model empirically it would be necessary to get information about aggregate consumption. To avoid the potential measurement problem, he develops an alternative model that replaces information about aggregate consumption with information about aggregate wealth, i.e., the market portfolio. This model is based on Constant Relative Risk Aversion (alternatively termed Constant Proportional Risk Aversion). (The models are separate because one cannot integrate a power function defined on positive wealth (consumption) with a normal distribution, which is defined on positive and negative wealth or consumption values.) Still more specific results are obtained by additionally assuming joint lognormality of securities being priced by the model and the return of the market portfolio. In addition, the paper is most likely the first to develop the idea of stochastic discount rates and to relate these to marginal utilities. Moreover, it points out that the stochastic discount rates come even more basically from the law of one price specialized so as not to allow riskless arbitrage opportunities. (As noted below, Ross (1977) works this latter result out exactly.)

Breeden (1979, Volume I, Chapter 5) develops the CCAPM in continuous time. As noted above, the model is particularly elegant in that it summarizes all the incentives to hedge shifts in consumption and portfolio opportunities with a one-factor relation between expected returns and consumption betas.

Campbell (1993, Volume I, Chapter 6) takes a different tack from Breeden, Merton, and Rubinstein using a loglinear approximation to the budget constraint to substitute out consumption from a standard intertemporal model. These models serve to show how much more complicated the consumption-investment decision is in a multiperiod setting.

To be more specific, intertemporal models are often based on time-separable power utility functions

\[ u(c_t) = \frac{1}{\gamma} c_t^{\gamma}, \quad \gamma < 1, \]

where \(1-\gamma\) is the coefficient of relative risk aversion and, as \(\gamma\) approaches zero, equation (15) approaches \(u(c_t) = \ln(c_t)\). Unlike their single-period brethren, the multiperiod models have both time-series and cross-sectional components. Following Cochrane (2001a), let \(r^*_i = \ln R_i\), \(r^*_t = \ln R_t\), \(\delta = e^{-\delta_t}\), and \(g\) be the expected growth rate in consumption, then the riskfree rate of interest is

\[ r^*_i = \delta_t + (1-\gamma)g - \frac{(1-\gamma)^2}{2}\sigma_i^2(\Delta \ln c). \]

Furthermore, Breeden (1979) and Campbell (1993) show risky securities are priced so that

\[ E(r^*_j) - r^*_j + \frac{\sigma_j^2}{2} = (1-\gamma)\sigma_{jc}, \]

\(j\)
where \( \sigma_{jc} \) is the covariance between the return on security \( j \) and consumption.

However, in what has been come to be known as the equity premium puzzle, equation (17) is incompatible with reasonable levels of risk aversion and with the observed volatility of consumption growth. For example, Cochrane (2001a, page 23) notes that over the last 50 years in the U.S., real stock returns have averaged 9% with a standard deviation of about 16%, while the real return on treasury bills has been about 1%. Thus, the historical annual market Sharpe ratio is about 0.5. Aggregate consumption growth has been about 1%. We can only reconcile these facts with the Hansen-Jagannathan bounds and equation (14) if the Pratt-Arrow measure of relative risk aversion is about 50. More surprising, Cochrane's (1996) estimate of relative risk aversion measure is on the order of 240!

Another way to see the problem is to assume that a representative investor has logarithmic utility. Then consumption would have to have an annual standard deviation of 50%.\(^2\) Even worse the (unbelievably high) risk aversion coefficients that fit equation (17) are completely the opposite of what we should find in fitting equation (16), which has come to be known as the riskfree rate puzzle. There are three possibilities. First, people are a lot more risk averse than almost anyone might have imagined. Second, historically the stock market returns were largely luck rather than an equilibrium compensation for risk. Third, something is fundamentally wrong with the model, including the use of aggregate consumption data.

Spurred by these observations, Campbell (1993) substitutes consumption out of the CCAPM. He obtains a pricing equation very close to Merton's (1973) ICAPM

\[
E(r^*_{jt}) - r^*_f + \frac{\sigma_{jm}^2}{2} = (1 - \gamma)\sigma_{jm} - \gamma \sigma_{jh},
\]

where \( \sigma_{jm} \) is the covariance of the return on security \( j \) with the return on the market portfolio and \( \sigma_{jh} \) is a (complicated) term that captures news about future returns on invested wealth. The Merton-Campbell result, as exemplified by equation (18), has important implications. Ignoring the variance term \( \sigma_j^2 \), the equation tells us that in a multiperiod model systematic risk \( \sigma_{jm} \) is not enough to price assets. We also have to worry about hedging against unfavorable shifts in the investment opportunity set.

If we assume that the single-period MV model approximates decisions made by power utility investors, then the reciprocal of the risk tolerance parameter is equal to the Pratt-Arrow measure of relative risk aversion \( (1 - \gamma) \). It is easy to estimate risk aversion as \( (1 - \gamma) = 1/T_m = (E(R_m) - R_f)/\sigma_m^2 \) using equation (4)—you can make the calculations on the back of an envelope. Cochrane's numbers mentioned a few paragraphs ago indicate that \( \gamma \approx -2 \) (or equivalently that the Pratt-Arrow measure of relative risk aversion is approximately equal to 3), which is much more reasonable than the consumption-based estimates. But, as equation (18) shows, this estimate ignores the possibility of shifts in investment opportunities. Campbell (1996), for example, finds that the estimated coefficient of relative risk aversion rises by a factor of ten or more if one allows for the empirically estimated degree of mean-reversion in postwar monthly U.S. data.
Furthermore, the $\sigma_{jk}$ term in equation (18), which captures news about changes in the opportunity set, helps to illustrate Cornell's (1981) point that it may be difficult to estimate consumption betas, as they would have to be inherently unstable to reflect the changes in investment opportunities.

The APT provides a series of insights into asset pricing not readily apparent in other asset pricing models. Rather than restricting tastes, the APT restricts beliefs. A factor model is assumed to describe returns, and the pricing equation follows from the absence of arbitrage. The APT's appeal is fourfold. First, it appears that we do not need to explicitly specify investor tastes. For example, Brennan (1981, page 353) characterizes the APT as a 'minimalist model of security market equilibrium' that is 'logically prior to our utility based models, and should be tested before the predictions of stronger utility specifications are considered.' Second, the assumption of a multifactor model expands the ideas of diversifiable and nondiversifiable risk, which as noted, Roll and Ross (1980) argue are the real intuition for the MV CAPM. Third, extending Roll and Ross' observation that a formal model may really front for a deeper but perhaps not fully formulated intuition, factor models provide a natural framework for thinking about how macroeconomic factors might influence the risk-return tradeoff. The appeal is obvious as other asset pricing models are sadly silent on how the real side of the economy affects asset pricing. Fourth, as Fama (1991) notes, factor models are an empiricist's dream as they fit very naturally into the regression framework employed in most empirical work.

The downside is that the factor structure offers only vague predictions about which variables are important in describing returns and expected returns. As we will see in section II, Brennan (1981, page 353) correctly foresees the problem (but not necessarily the intensity of the debate) this may create: 'I am concerned lest the practical influence of the APT on our research programme should be to lead it in the direction of ad hoc data manipulation and atheoretical correlation analysis.'

To be more specific, the APT assumes returns conform to a $k$-factor linear model

$$r_j = E_j + \beta_j \delta_1 + \ldots + \beta_{jk} \delta_k + \epsilon_j, \quad j = 1, \ldots, n,$$

(19)

where $r_j$ and $E_j$ are the rate of return and expected rate of return on asset $j$, respectively, the $\delta_k$ are zero-mean common factors, and the $\epsilon_j$ are zero-mean asset specific disturbances assumed to be uncorrelated with the $\delta_k$ and with each other. The fundamental assumption is that the number of factors is much smaller than the number of assets. In matrix notation

$$R = E + B \delta + \epsilon,$$

(20)

where $R$, $E$, and $\epsilon$ are $n$-vectors, $B$ is an $n \times k$ matrix, and $\delta$ is a $k$-vector. Ross (1976, 1977) shows that in the absence of riskless arbitrage this implies the existence of a constant $\gamma_0$ and a $k$-vector $\gamma_1$ such that

$$E \approx \gamma_0 t + B \gamma_1.$$

(21)

The key result is that only the systematic (nondiversifiable) risks are priced. The result is intuitively pleasing and it is developed under a set of very weak conditions—no restrictions have been placed on investors' tastes except that more is preferred to less. Furthermore, there is no need to observe the market portfolio.
However, Shanken (1982) makes the case that the APT is not easily testable unless equation (21) holds as equality

\[ E = \gamma_0 + B\gamma_1. \]  

Ironically, as noted, Conner (1984) obtains an exact pricing relation by relying on the principles of competitive equilibrium rather than on an arbitrage technique. So tastes work their way back into the model. Moreover, a crucial condition in his derivation is the role played by the market portfolio. In Connor's model, the idiosyncratic risk, defined relative to a given factor structure, is completely diversified away in the market portfolio.

But, the issue of whether exact factor pricing is required for empirical testing is open for debate—a debate, which we explore in detail in section VII.

Ross (1977, Volume I, Chapter 7) develops the APT relating it to the mean-variance and state-preference approaches to asset pricing. Quite aside from an intuitive derivation of the APT in a return-generating framework, the paper contains a wealth of other insights. For example, on page 190 he dismisses the assumptions of quadratic utility or normality upon which the MV model is often based: 'Neither of these assumptions, however, is particularly appealing on intuitive economic grounds; normality has only an inappropriate (and careless) application of the central limit theorem to recommend it and quadratic utility functions are implausible for any agents, let alone for all of them.' In section 9.3, he presents one of the earliest demonstrations that the absence of arbitrage is equivalent to the existence of market equilibrium with positive prices.

Before proceeding to the empirical work, I note that any survey of asset pricing would be incomplete without mentioning parameter-preference valuation, a model born in the 1970s, but one that attracted little empirical interest until the turn of this century. So let's try one last intuition for mean-variance decision-making. In a mean-variance economy, investors balance two moments—expected return and variance—in forming their portfolios. The mean-variance pricing equation links expected return and covariance risk—the contribution a security makes to the market portfolio's variance. Rubinstein (1973) takes a Taylor series expansion of the expected utility problem to show that in general an investor may care about an infinite sum of moments in forming his or her portfolio. The resulting pricing equation then links expected returns to an infinite sum of co-moments. Special cases include quadratic utility, which leads to the mean-variance model, and (separable) cubic utility, which leads to a mean-variance-skewness model. Kraus and Litzenberger (1976) also focus on a skewness model and like Rubinstein show that the pricing equation links expected returns to covariance risk—the contribution a security makes to the market portfolio's variance—and co-skewness—the contribution a security makes to the market portfolio's skewness. The skewness model attracted little empirical interest after Kraus and Litzenberger's test of it. That is until Harvey and Siddique (2000) tested a conditional form of the model. Their paper is of interest not only because of the empirical work, but also because they develop the model as a special case of the stochastic discount methodology.

With the theories in place, we turn to the empirical tests.
II Pre-1990 Tests of the Mean-Variance Capital Asset Pricing Model

For over thirty-five years, MV CAPM has been one of the central paradigms of financial economics. As noted, from a theoretical point of view it represents an almost perfect blend of elegance and simplicity. Moreover, from an empirical point of view the model appears to be readily testable. Betas are easily estimated from standard time series regressions. And, a linear risk-return tradeoff seems to be tailor-made for empirical testing. But is it? In this section, we review the extensive literature that tests the CAPM together with the body of literature, which asserts that tests of the CAPM are ambiguous at best.

Each of the following seven statements has an implication for how we might judge whether the CAPM is true or false. First, the market portfolio is MV efficient. Second, there is a linear relation between the expected returns and market betas of securities, i.e., securities plot on the SML. Third, market betas are the only measures of risk needed to explain the cross-section of expected returns. Fourth, in the riskfree asset version of the model, the market portfolio is the tangency portfolio, i.e., the point of tangency between a ray emanating from the riskfree interest rate and the minimum-variance frontier of risky assets. Fifth, in the riskfree asset version of the model, the separation property implies that all investors will hold some combination of the riskfree asset and the market portfolio. Sixth, there is at least one positively weighted efficient portfolio. Seventh, the Invariance Law of Prices implies that if the CAPM determines prices and means, variances, and covariances are constant over time, then there is an exact linear relation between the values of any three assets.

There is disagreement, however, over which of the implications we should test, with some of the implications being dismissed out of hand. For example, Huang and Litzenberger (1988, pages 301-302) state: 'A strong prediction of the CAPM is two fund separation. … While to the best of our knowledge no researcher takes such a prediction seriously, another strong prediction that the market portfolio is on the portfolio frontier has been subjected to extensive testing.' Furthermore, Turnbull and Winter (1982) and Gibbons and Ferson (1985) comment that Cheng and Grauer's (1980) Invariance Law can be tested simply by observing three securities at three points in time. But times change. Lo and Wang (2000) examine the implications of portfolio theory and two-fund separation for the cross-sectional behavior of equity trading volume. First, they show that if two-fund separation holds, share turnover must be identical for all securities. If \((K+1)\)-fund separation holds, share turnover satisfies an approximate linear \(K\)-factor structure. Then, they present strong evidence against two-fund separation as well as evidence that a two-factor linear model drives volume.

The statement that there is at least one positively weighted efficient portfolio is a necessary condition for the CAPM to be true, but it is not, of course, sufficient. While it has not attracted much attention in the literature focusing on tests of positive risk – return tradeoffs, it may be worth thinking about given the extreme sensitivity of mean-variance portfolio weights to small perturbations in the means documented in Best and Grauer (1991, 1992) and Grauer (1999). The calculations in these papers indicate that even the slightest perturbation in an equilibrium set of expected returns would mean that there would be no positively weighted portfolios on the minimum-variance frontier. So, without almost exact security market line pricing, markets may not clear in a mean-variance economy.

Be that as it may, it is the first three statements, especially the second and third, which have formed the basis for most empirical tests of the MV model. In one sense, this is perfectly
understandable, as a positive risk-return tradeoff has fascinated economists since long before the advent of the CAPM. In another sense it is disturbing that the SML receives the bulk of attention, as it requires additional sets of assumptions if it is to be testable, and it may not be testable at all.

One of the most basic problems arises from the fact that the SML is an *ex ante* relationship relating expected returns to betas, but all we can observe is a time series of *ex post* returns. Clearly, any theory that purports to be testable must establish a relationship between *ex post* returns and *ex ante* beliefs. Fama and MacBeth (1973), Fama (1976), and Ross (1978) argue that the most straightforward way to accomplish this is to require strong homogeneity of individual beliefs. But, as Ross (1978, page 889) observes: 'Of course, casual observation suggests quite the opposite; it seems clear that investors do not show the homogeneity of beliefs which characterize our theories. To the contrary, it is widely believed that it is differences in beliefs and disparate changes in these beliefs which occasion trade …'.

Thus, although the MV CAPM is a single-period model it seems clear that the theory and empirical tests envision a multiperiod setting where, period-by-period, investors choose MV-efficient portfolios. In this multiperiod scenario, there are no compelling theoretical reasons for assuming that the opportunity set is constant. However, the early tests were more or less forced to assume a constant mean vector and covariance matrix. See, for example, Fama (1976, pages 272-273 and page 344). Miller and Scholes (1972, page 73) sum it up rather nicely: 'Expectations, in short, are assumed to be realized at least "on the whole and on the average in the long run," even though the returns for any particular firm or time period may have a large "unanticipated" component.'

These assumptions may be right, but even casual observations suggest they may not be. As Elton (1999, page 1199) points out:

There are periods longer than 10 years during which stock market realized returns are on average less than the risk-free rate (1973-1984). There are periods longer than 50 years in which risky long-term bonds on average underperform the risk free rate (1927 to 1981). Having a risky asset with an expected return above the riskless rate is an extremely weak condition for realized returns to be an appropriate proxy for expected returns, and 11 and 50 years is an awfully long time for such a weak condition not to be satisfied. In the recent past the United States has had stock market returns of higher than 30 percent per year while Asian markets have had negative returns. Does anyone honestly believe that this is because this was the riskiest period in history for the United States and the safest for Asia?

Of course, Elton's points are debatable. One could equally well argue that the relationship between *ex ante* expectations and *ex post* returns depends on the volatilities of the asset classes. For example, given such noise in equities, there must be quite a few decade-long periods when they underperform the "riskless" rate within an infinitely long time series. But, his points are worth thinking about.

Even so, at a more fundamental level, the CAPM itself may rule out the possibility of *ex post* returns being drawn from *ex ante* return distributions. This possibility is discussed below.

As the *ex ante* variables are unobservable, it would then seem that the most natural way to test the model is by regressing *ex post* average excess returns on betas

18
\[ \bar{r}_j - \bar{r}_j = \gamma_0 + \gamma_1 \beta_j + e_j, \]  

where \( \bar{r}_j \) and \( \bar{r}_j \) are time-series average rates of return on risky asset \( j \) and the riskfree asset, respectively, and the \( \beta_j \)'s are estimated from a time-series regression

\[ r_{jt} = \alpha_j + \beta_j r_{mt} + e_{jt}, \text{ for each asset } j, \]  

where \( r_{mt} \) is the rate of return on the market portfolio at time \( t \). Unfortunately, there are three problems associated with this two-pass cross-sectional approach. The most fundamental problem of whether \textit{ex post} returns reflect \textit{ex ante} expectations, and the closely related problem of how to reconcile stationary return distributions with changing riskfree rates of interest and betas, are discussed below. The third problem is that the error terms are almost certainly heteroscedastic and cross-sectionally correlated.

In an effort to get around the latter problem, Fama and MacBeth (1973) propose the following heuristic method that has been employed in numerous subsequent studies. In each month, they run the cross-sectional regression

\[ r_{jt} - r_{jt} = \gamma_{0t} + \gamma_{1t} \beta_{jt} + e_{jt}, \]  

where the \( \beta_{jt} \)'s are sometimes estimated once a year based on a five-year moving window and sometimes are simply the full-period betas, i.e., \( \beta_{jt} = \beta_j \) is estimated from the full-period time-series regression (24). (Equation (25) is also run in raw return form. That is, without subtracting the riskfree rate of interest from the right-hand side.) Tests of the model are based on the time series of \( \gamma_{0t} \) and \( \gamma_{1t} \). In most cases the betas are estimated from monthly data. However, Kothari, Shanken, and Sloan (1995) also use betas estimated from annual data. The \( \gamma \)'s have interesting interpretations: \( \gamma_{0t} \) is the excess return on a zero-beta portfolio, and \( \gamma_{1t} \) is a zero-weight portfolio whose return is the return on the market in excess of the zero-beta rate. It can be shown that if \( \beta_{jt} = \beta_j \), then the averages of the time series of the \( \gamma_{0t} \)'s and \( \gamma_{1t} \)'s are equal to \( \gamma_0 \) and \( \gamma_1 \) estimated from equation (23). Fama and MacBeth (1973) suggest using the standard deviations of \( \gamma_{0t} \) and \( \gamma_{1t} \) to generate the sampling errors of \( \gamma_0 \) and \( \gamma_1 \). The tests are performed as simple \( t \)-tests

\[ t(\bar{y}_j) = \frac{\bar{y}_j}{s(\gamma_j)/\sqrt{T}}, \]  

where \( \bar{y}_j \) and \( s(\gamma_j) \) are the average and standard deviation of the \( \gamma_{jt} \)'s and \( T \) is the number of time series observations.

Black, Jensen, and Scholes (1972) take a different approach suggesting a time-series test based on the regression

\[ r_{jt} - r_{jt} = \alpha_j + \beta_j (r_{mt} - r_{jt}) + e_{jt}, \text{ for each asset } j. \]  

The intercept \( \alpha_j \) measures the deviation of security \( j \)'s mean from the SML. They test the null hypothesis that \( \alpha_j = 0 \) for each security \( j \).
The multivariate tests of Gibbons, Ross, and Shanken (1989, Volume I, Chapter 8), Jobson and Korkie (1982, 1985) and MacKinlay (1987) are important methodological refinements. (For added detail, see Campbell, Lo and MacKinlay (1997) and Stewart (1995, 1997).) In order to develop the multivariate tests, let $\alpha = (\alpha_1, \ldots, \alpha_n)'$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ be $n$-vectors containing the intercepts and error terms from Black, Jensen, and Scholes' time series regression, equation (26). Assume $E(\epsilon_i) = 0$, $E(\epsilon_i \epsilon_j') = \Sigma$, and $\text{cov}(r_{mt}, \epsilon_i) = 0$. Furthermore, assume the $\epsilon_i$ are jointly normally distributed. Then, Gibbons, Ross and Shanken (1989) show that we can jointly test the hypothesis that $\alpha = 0$ using

$$J = \frac{(T - N - 1)}{N} \left( 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right) \hat{\epsilon}' \hat{\Sigma}^{-1} \hat{\epsilon},$$

(27)

where $\hat{\mu}_m$ and $\hat{\sigma}_m$ are the average excess return and standard deviation of the 'market' portfolio. Under the null hypothesis, $J$ is unconditionally distributed central $F$ with $N$ degrees of freedom in the numerator and $T-N-1$ degrees of freedom in the denominator. It is noteworthy that this test is exact for finite samples. It does not rely on asymptotic theory.

A useful economic interpretation can be made of the test statistic using the efficient set mathematics. Gibbons, Ross, and Shanken show that

$$J = \frac{(T - N - 1)}{N} \left( 1 + \frac{\hat{\mu}_q^2}{\hat{\sigma}_q^2} \right) \left( 1 + \frac{\hat{\mu}_m^2}{\hat{\sigma}_m^2} \right),$$

(28)

where the portfolio denoted by $q$ represents the ex post tangency portfolio constructed from the $N$ included assets plus the market portfolio. Thus, the test can be interpreted either as a multivariate test of the securities' deviations from the SML in mean - beta space (see equation (27)) or as the difference in the squared Sharpe ratios of the proxy and tangency portfolios in mean - standard deviation space ($\epsilon$ see equation (28)).

It is important to note, however, that the test requires the use of a proxy for the market portfolio. Moreover, it assumes stationary return distributions and, implicitly, a constant riskfree interest rate. I find it troubling to claim that one is estimating the tangency portfolio in periods like the last thirty years where the riskfree interest rate so obviously did change. This is especially so in light of the extreme sensitivity of MV portfolio weights and the position of the tangency portfolio to small perturbations in the means and the riskfree rate of interest documented in Best and Grauer (1991, 1992) and Grauer (1999).

Jensen (1972, Volume I, Chapter 9) provides an excellent overview of the theory and early tests of the model. As he reports, in the first direct test of the MV CAPM, Douglas (1969) finds the estimated slope of the security market line is too flat and the intercept is too large. In an influential paper, Miller and Scholes (1972) replicate the Douglas study and provide a detailed analysis of the possible econometric difficulties involved in estimating the relationship. They conclude that measurement error in the betas seems to contribute to the Douglas result. That is, with any two-variable regression, measurement error in the independent variable can cause the estimated slope to be flatter than the true slope. Shortly thereafter, Black, Jensen, and Scholes
(1972), Blume and Friend (1973), and Fama and MacBeth (1973) produce the first extensive tests of the model. They focus on the cross-sectional expected return - beta tradeoff and the special prediction of the Sharpe-Lintner version of the model that the returns on 'zero-beta' portfolios have expected returns equal to the riskfree rate of interest. They group securities into portfolios to reduce the measurement errors in the betas. Their well-known findings show that the average return - beta plot is almost linear, but the estimated slope of the SML is too flat and the intercept is too high. The evidence is interpreted as providing grounds for rejection of the original Sharpe-Lintner model and as being consistent with Black's (1972) zero-beta CAPM.

In addition to discussing the early tests, Jensen reviews ways in which the assumptions underlying the model might be relaxed. First, he discusses the underlying assumptions necessary to justify the single-period of terminal wealth assumption, see Fama (1970). Second, he discusses how to relax the assumption of riskless borrowing and lending opportunities, see Black (1972) and Brennan (1971). Third, he discusses the effect of the existence of nonmarketable assets, see Mayers (1972). Fourth, he discusses the effect of differential taxes on dividends and capital gains, see Brennan (1970). Fifth, he provides what I believe is the first test of Merton's (1973) ICAPM.

As we shall see, there have been many subsequent rejections of the model and a number of interesting innovations in the econometrics of testing the model. But, in my view, the most fundamental doubts about the empirical evidence arise from logical rather than statistical considerations. Perhaps surprisingly, most of the arguments are made in a single-period framework and are illustrated with fairly simple numerical examples.

Roll's (1977, Volume I, Chapter 10) critique is the best known. The paper provides an excellent review of the efficient set mathematics as well as a detailed critique of the early tests. More importantly, Roll is the first to appreciate the significance of the statement: 'securities plot on the SML if and only if the market portfolio is MV efficient'. The theory asserts that a particular portfolio, the market portfolio, is MV efficient. Thus, he asserts that the theory is not testable (at least in terms of the SML) unless that portfolio is observable and used in the tests. The problem is not a measurement error problem in the usual sense of a "sampling problem". It is more fundamental: we cannot measure the true betas. As Roll notes, betas are a function of the portfolio from which they are calculated. Thus, if the market portfolio betas could be observed, they might well be different from, say the betas of an equal-weighted portfolio or from the betas of any portfolio defined on some subset of the assets.

On the other hand, Fama (1991, page 1590), states: 'Stambaugh's (1982) evidence that tests of the SLB model are not sensitive to the proxy used for the market suggests that Roll's criticism is too strong, but this issue can never be entirely resolved.' In a pragmatic, if somewhat limited, sense Fama is right—if we interpret the CAPM as simply providing a description of the pricing of U.S. securities. Unfortunately, that interpretation and Stambaugh's evidence does not directly address Roll's concern. I don't find Stambaugh's argument convincing for this reason and in light of related issues, raised in Grauer (1978, 1999), Roll and Ross (1994), and Kandel and Stambaugh (1995), which are discussed below.

Cheng and Grauer (1980, Volume I, Chapter 11) identify further ambiguities associated with the tests. As noted above, the early tests were more or less forced to assume a constant mean vector and covariance matrix. Moreover, by using either equation (24) or (26) to estimate betas, researchers implicitly assume that the betas are constant as well. Thus, early tests of the CAPM
are really tests of a joint hypothesis: the CAPM determines prices, and means, variances, covariances and betas are constant over time. Unfortunately, the joint hypothesis leads to a remarkable and rather unpalatable conclusion: relative prices never change. (Neither does the riskfree rate of interest.) Cheng and Grauer present an alternative test that focuses on the Invariance Law of Prices, which asserts that if the CAPM determines prices, and means, variances, and covariances are constant over time, then there is an exact linear relation between the values of any three assets. Thus, they test an admittedly less intuitively appealing implication of the CAPM under a weaker set of assumptions (where betas and the riskfree rate are not assumed to be constant), without having to identify the 'true' market portfolio, and without the use of \textit{ex post} return data. Perhaps not surprisingly, this alternative test of the Invariance Law provides little support for the CAPM.

Turnbull and Winter (1982) and Sweeney (1982) assert that the assumption of stationarity leads to an internal inconsistency. Furthermore, they argue that relaxing the stationarity assumption will alleviate the problem. Cheng and Grauer (1982, Volume I, Chapter 12) maintain that the problem is more fundamental: when the single-period CAPM is embedded in a multi-period setting, \textit{ex post} returns are not drawn from the \textit{ex ante} return distributions envisioned by investors. Specifically, they show that if return distributions and the riskfree rate do not change over time, then at any date all risky securities realize the same rate of return. If the riskless rate changes over time, securities realize different rates of return, but \textit{ex post} returns are not drawn from \textit{ex ante} return distributions. Finally, they show that the lack of correspondence between \textit{ex post} and \textit{ex ante} returns cannot be corrected by assuming nonstationary return distributions. While finance may be blessed with the cleanest and most plentiful data in all of economics, it is of little comfort to those attempting to test the CAPM using the SML if these \textit{ex post} returns do not reflect investors' \textit{ex ante} beliefs.

This result is both fundamental and controversial. Surprisingly, I have not seen it even mentioned in any subsequent tests other than Gibbons and Ferson (1985). The result appears to be robust, however. Rosenberg and Ohlson (1976, Volume I, Chapter 13) express similar concerns in a continuous-time setting. They show that the joint properties of identically distributed security rates of return over time and portfolio separation at each date, which characterizes the simpler version of Merton's (1973) continuous-time model, force all risky securities in equilibrium at any date to have the same rate of return. While this "equal rate of return for all securities" result arises in both continuous time and a period-by-period application of the single-period CAPM, it is best not to underestimate the controversy it generates. For example, Merton's (1975, page 665) response to Rosenberg and Ohlson, which was still in working paper form at that point in time, is short if not sweet: 'Of course, this is nonsense.'

III Post-1990 Tests of the Mean-Variance Capital Asset Pricing Model: Anomalies and Fama and French's Three-Factor Model

In the 1990's the third implication of the model, that market betas are the only measures of risk needed to explain the cross-section of expected returns, has created by far the most attention. The literature surrounding tests of this proposition is fascinating. The tests set off a multifaceted debate concerned with the efficacy of the testing methodology, on the one hand, and the interpretation of the results, on the other. It is not clear whether the tests have meaning, for example, or even whether they are tests of the MV CAPM. Fama and French (1993, 1995, 1996a, 1996b, 1997, 1998) argue that the empirical successes of their three-factor model suggest that it is an equilibrium asset-pricing model, a three-factor version of the ICAPM or the APT.
Moreover, the tests continue to blur the distinction between tests of market efficiency and tests of asset-pricing models. For example, we see interpretations of cross-sectional tests of asset pricing models drawn from models of irrational behavior that first gained attention in predicting the time series of returns. In addition, we see conditional tests of asset pricing models, where the conditioning variables are drawn from the same return predictability literature.

The reader might be surprised to see Fama's second classic review of market efficiency(1991, Volume I, Chapter 14) in a volume purporting to review tests of asset pricing models. However, the two literatures are intimately related by a joint hypothesis. As Fama (1991, page 1589) explains:

empirical research on asset-pricing models … does not place itself in the realm of tests of market efficiency, but this just means that efficiency is a maintained hypothesis. Depending on the emphasis desired, one can say that efficiency must be tested conditional on an asset-pricing model or that asset-pricing models are tested conditional on efficiency. … Moreover, many of the front-line empirical anomalies in finance (like the size effect) come out of tests directed at asset-pricing models. Given the joint hypothesis problem, one can't tell whether such anomalies result from misspecified asset-pricing models or market inefficiency.

As such, the article provides an excellent overview of the empirical work on efficient markets and asset-pricing tests from the early 1970's through the early 1990's.

An extensive literature documents deviations from the linear risk-return tradeoff that are related to firm-specific variables such as size (Banz (1981)), earnings yield (Basu (1977, 1983)), and the ratio book-to-market equity (Rosenberg, Reid and Lanstein (1985)). Similarly, DeBondt and Thaler (1985) show that stocks that exhibit extreme behavior either as winners or losers over one to five year periods in the past display mean-reverting behavior over the next five years. That is, past winners (losers) become future losers (winners). In contrast, Jegadeesh and Titman (1993) find that forming portfolios based on short-term past returns leads to a continuation of short-term returns. The past losers tend to be future losers and past winners are future winners.11 Because the CAPM apparently cannot explain these patterns of average returns, they are called anomalies.

If this anomalies-based literature did not hold center stage earlier, it certainly has since Fama and French (1992, Volume I, Chapter 15). They report that two easily-measured variables, size and book-to-market equity, combine to capture the cross-sectional variation in average stock returns associated with market beta, size, leverage, book-to-market equity, and earnings-price ratios. Moreover, in a shot straight at the heart of the CAPM, the relation between market beta and average return is flat, if the tests allow for variation in beta that is unrelated to size. This occurs even when beta is the only explanatory variable.12 The empirical tests patterned after Fama and MacBeth (1973) incorporate two innovations. As noted, securities are grouped into portfolios that are double-sorted on size and beta. Then, a group beta is assigned to the individual stocks in a group, and the cross-sectional regressions are run on individual stocks rather than on portfolios.13

The reaction to Fama and French's (1992) paper has been swift and far-reaching. Some believe the CAPM is dead. Others await a meaningful test. Some believe a three-factor model that is consistent with rational behavior has replaced the CAPM. Others believe that taxes and liquidity are missing factors in a rationally based CAPM. Still others believe the three-factor model is consistent with irrational behavior.

One school argues that the CAPM may be spuriously rejected. However, the arguments are wide-ranging. Lo and MacKinlay (1990), Black (1993), and MacKinlay (1995) maintain that the CAPM anomalies may be the result of data snooping. (Ironically, as noted in section I, Fama (1991, page 1595) warns against this very problem.) Kothari, Shanken and Sloan (1995) argue that there is survivorship bias in the returns used to test the model, especially in terms of the book-to-market returns. Roll and Ross (1994), Kandel and Stambaugh (1995), and Grauer (1999) suggest that the cross-sectional tests are susceptible to the use of poor proxies for the market portfolio. For example, using population parameters Roll and Ross show a proxy can be nearly MV efficient and yield a zero slope in a regression of means on betas (calculated relative to the proxy). Finally, Kan and Zhang (1999a) highlight a surprising statistical property of the standard two-pass cross-sectional regression methodology: a factor uncorrelated with asset returns will appear to be priced with high probability.\(^{14}\)

A second school of thought, championed by Fama and French (1993, 1995, 1996a, 1996b, 1997, 1998), Fama (1996), and Davis, Fama, and French (2000), argues that asset pricing is rational and conforms to a three-factor ICAPM or APT that does not reduce to the CAPM. That is, the high returns earned by value stocks (those with low prices relative to earnings, dividends, book assets or other measures of fundamental value) arise because these stocks are fundamentally riskier than growth or glamour stocks.

A third school contends that taxes and liquidity are missing factors in a rationally-based CAPM. For example, Klein (2001) introduces a capital gain lock-in effect, while Amihud and Mendelson (1986), Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), Amihud (2000), and Pastor and Stambaugh (2001) make the case for higher expected returns in illiquid stocks.

A fourth school suggests that it is investor irrationality that prevents the three-factor model from collapsing to the CAPM. Specifically, irrational pricing causes the high premium for relative distress. Proponents of this view include Lakonishok, Shleifer, and Vishny (1994), Haugen (1995), MacKinlay (1995), and Daniel and Titman (1997).

Consider first the school that suggests the CAPM may be spuriously rejected. Black (1993) decries the fact that the Fama-French model is based on data mining—that is, Fama and French employ size and book-to-market equity variables that previous researchers had shown to be anomalous—and not on theory. Lo and MacKinlay (1990, Volume II, Chapter 1) examine the data snooping biases in tests of asset pricing models. They argue that tests of an asset-pricing model may yield misleading inferences when properties of the data are used to construct the test statistics. For example, the tests are often based on returns to portfolios of common stocks, where portfolios are constructed by sorting on some empirically motivated characteristic such as size (the market value of equity). They use analytical calculations, Monte Carlo simulations, and
two empirical examples to show that the effects of data snooping can be substantial. However, given that it is difficult to specify the adjustment that should be made for data snooping, the main message is that the biases should at least be considered as a potential explanation for deviations from the model.\textsuperscript{15}

MacKinlay (1995, Volume II, Chapter 2) identifies the two possible alternatives to the MV CAPM: risk-based multifactor asset pricing models and nonrisk-based models that address biases in empirical methodology or the presence of irrational investors. He develops a framework which shows that, \textit{ex ante}, CAPM deviations due to missing risk factors will be very difficult to detect empirically, whereas deviations resulting from nonrisk-based sources are easily detectable. He then argues that the results suggest that multifactor pricing models alone do not entirely resolve CAPM deviations.

Kothari, Shanken and Sloan (1995, Volume II, Chapter 3) maintain that Fama and French's results are influenced by a combination of survivorship bias in the COMPUSTAT database affecting the high book-to-market stocks' performance and period-specific performance of both low book-to-market past winner stocks, and high book-to-market past loser stocks. They also point out that given the low power of the tests of a positive market risk premium, the Fama-French evidence provides little basis for rejecting the null hypothesis of a nontrivial six percent per annum risk premium over the post-1940 period.

The regression tests focus on whether beta is the only measure of risk needed to explain the cross-section of expected returns. Roll and Ross (1994), Kandel and Stambaugh (1995), and Grauer (1999) highlight the dangers of focusing exclusively on mean-beta space. Roll and Ross (1994, Volume II, Chapter 4) categorize proxies that produce particular relations between expected returns and true betas. It is particularly striking that a market proxy can be almost MV efficient even though the slope from an ordinary least squares (OLS) regression of population expected returns on population betas is zero. Conversely, Kandel and Stambaugh (1995, Volume II, Chapter 5) show that there can be a near perfect OLS fit between means and betas calculated relative to a proxy that is grossly inefficient. More importantly, they show that, in a generalized least squares (GLS) regression of mean returns on betas, the slope and R-square are determined uniquely by the mean-variance location of the market index relative to the minimum-variance boundary. However, neither Roll and Ross nor Kandel and Stambaugh verify whether the minimum-variance frontier contains a positively weighted portfolio. Thus, we cannot be sure whether their results hold if the CAPM is true, i.e., if the positively weighted market portfolio is MV efficient. Furthermore, any reasonable proxy portfolio should contain positive weights. Roll and Ross are able to construct one example where a proxy portfolio contains positive weights, but the proxies in Kandel and Stambaugh's paper do not contain positive weights.

Grauer (1999, Volume II, Chapter 6) overcomes these difficulties by examining scenarios where the MV CAPM is true and where it is false. When the CAPM is true, the implications of the model mentioned at the beginning of section II hold exactly. In this case, the findings are not as dramatic as those reported by Roll and Ross and Kandel and Stambaugh. For example, the slopes of OLS and GLS regressions, that employ betas calculated from almost-efficient positively weighted proxy portfolios, are too flat—but not equal to zero. On the other hand, when the CAPM is false and the market portfolio is almost efficient, the slope of a regression of means on market betas can be zero. Perhaps more important, the coefficients of OLS, GLS, or both OLS and GLS regressions that employ market betas can take on exactly the same values as when the model is true, even though the market is grossly inefficient. The lack of any clear-cut
agreement among the different implications of the CAPM when it is false is particularly disturbing. It does not bode well for those seeking to design an unambiguous test of the model.

In a related study, Kan and Zhang (1999a, Volume II, Chapter 7) investigate the statistical properties of the standard two-pass cross-sectional tests of asset pricing models with misspecified factors. More specifically, they study the extreme case of a useless factor that is independent of all the asset returns. Simulation evidence suggests that the traditional $t$-test rejects a zero risk premium for a useless factor with a probability more than twice the size of the test for a typical length of time series used in empirical studies. Surprisingly, the problem is exacerbated when the number of time series observations increases. Kan and Zhang (1999b) extend the analysis examining generalized method of moments (GMM) tests of stochastic discount factor models with useless factors. The GMM tests fare no better. In some cases a misspecified model with a useless factor is more likely to be accepted than the true model.

The second school, championed by Fama and French, argues that asset pricing is rational and conforms to a three-factor risk-based ICAPM or APT that does not reduce to the CAPM. Fama and French (1993) extend the analysis of Fama and French (1992) by providing a three-factor model of asset pricing. The model says the expected return on a security in excess of the riskfree rate is explained by the sensitivity of its return to three factors. The first is the excess return on the market portfolio. The second is the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks (SMB, small minus big). The third is the difference between the return on a portfolio of high-book-to-market stocks and the return on a portfolio of low-book-to-market stocks (HML, high minus low). Their subsequent empirical work attempts to establish whether the factors have economic meaning: specifically that HML proxies for financial distress. For example, Fama and French (1995) show that there are common factors in fundamentals like earnings and sales that look a lot like the SMB and HML returns. Moreover, Fama and French (1998) provide out-of-sample international evidence for the premium earned by value (high book-to-market) over growth (low book-to-market) stocks.

Fama and French (1996a, Volume I, Chapter 16) show that, except for the continuation of short-term returns, the CAPM anomalies largely disappear in a three-factor model. The article makes two important contributions. First, as noted, Fama and French (1992) employ a twist on the Fama and MacBeth (1973) cross-sectional test methodology using individual stock data. By way of contrast, Fama and French (1996a) employ univariate time-series tests similar to those of Black, Jensen and Scholes (1972) and multivariate time-series tests similar to those of Gibbons, Ross and Shanken (1989) using grouped data. Thus, the results appear to be robust to the exact econometric method employed. Second, the paper provides an excellent discussion of various interpretations of the results. Fama and French (1996a, page 76) acknowledge the minimalist interpretation of their results: ‘that is, we have simply found three portfolios that absorb most of the anomalies of the CAPM. In other words, without knowing why, we have stumbled on explanatory portfolios that are close to three-factor MMV.’ (MMV their shorthand for multifactor-minimum variance.) Not surprisingly, the rest of the article includes a spirited defense of an ICAPM or APT interpretation of the results as well as a strong denunciation of alternate positions.

Loughran (1997), however, observes a contradiction: although the finance literature declares that value firms have reliably higher realized returns than growth firms, the realized returns on value and growth money managers are not materially different. The paper explores book-to-market characteristics across the dimensions of firm size, exchange listing, and calendar
seasonality. The results may be summarized as follows. 'Fama and French's findings are driven
by two features of the data: a January seasonal in the book-to-market effect, and exceptionally
low returns on small, young, growth stocks. In the largest quintile of all firms (accounting for
73% of the total market value of all publicly traded firms), book-to-market has no significant
explanatory power on the cross-section of realized returns during the 1963-1995 period. Thus,
book-to-market as such would have less importance to money managers than the literature would
have led us to believe.' Loughran (1997, page 249).

The third school argues that taxes and liquidity may be missing factors in a rationally based
asset pricing model. As noted earlier, Brennan (1970) first examined the effect of differential
taxes on dividends and capital gains on the CAPM. Klein (2001) extends the model to allow for
the presence of a capital gain lock-in effect. Empirical tests on the cross-section of stock returns
find that long horizon return reversal is primarily attributable to the effect of investors' accrued
capital gains. Moreover, the tests are robust to the addition of size and past total return.

Other papers allege a role for liquidity in asset pricing. Amihud and Mendelson (1986)
argue for higher expected returns in illiquid stocks, though they do not suggest this is a risk
premium. It is more like a credit spread on a corporate bond. Nonetheless, they find a relation
between expected returns and liquidity after adjusting for risk using the CAPM. Brennan and
Subrahmanyam (1996) find a significant relation between required rates of return and liquidity
measures after adjusting for risk using the Fama and French's three factors and the effects of the
stock price level. Brennan, Chordia, and Subrahmanyam (1998) examine the relation between
stock returns, measures of risk and several non-risk security characteristics, including the book-
to-market ratio, firm size, stock price, dividend yields, and lagged returns. They attempt to
determine whether non-risk characteristics have marginal explanatory power relative to the
arbitrage pricing theory. They employ two risk adjustments based on Connor and Korajczyk's
(1988) principal components approach and Fama and French's (1993) characteristic based three-
factor approach. Regardless of the method used to adjust for risk, they find a strong negative
relation between average returns and trading volume, which is consistent with a liquidity
indicate that the liquidity area has a lot of promise.

The fourth school, exemplified by papers by Lakonishok, Shleifer, and Vishny (1994,
Volume II, Chapter 8), Haugen (1995), MacKinlay (1995), and Daniel and Titman (1997),
argues that it is investor irrationality that prevents the three-factor model from collapsing to the
CAPM. Daniel and Titman (1997) suggest that the value-premium traces to the value
characteristic and not to risk. The basic story is that investors like growth stocks (strong firms)
and dislike value stocks (weak firms). The result is a value premium that is not due to risk.

Lakonishok, Shleifer, and Vishny (1994) add investor overreaction to the story. The basic
idea is that contrarian investment strategies that buy value stocks can exploit the irrational
behavior of investors attracted to glamour stocks. Lakonishok, Shleifer, and Vishny (1994, page
1542) pick up the story: 'While there is some agreement that these [value] strategies have
produced superior returns, the interpretation of why they have done so is more controversial.
Value strategies might produce higher returns because they are contrarian to "naive" strategies
followed by other investors. These naïve strategies might range from extrapolating past earnings
growth too far into the future, or to assuming a trend in stock prices, to overreaction to good or
bad news, or to simply equating a good investment with a well-run company irrespective of
price. Regardless of the reason, some investors tend to get overly excited about stocks that have
done very well in the past and buy them up, so that these "glamour" stocks become overpriced. Similarly, they overreact to stocks that have done very badly, oversell them, and these out-of-favor "value" stocks become underpriced. Contrarian investors bet against such naïve investors. Because contrarian strategies invest disproportionately in stocks that are underpriced and underinvest in stocks that are overpriced, they outperform the market. Clearly this view stands in stark contrast to the Fama-French position that value stocks are fundamentally riskier.

Lakonishok, Shleifer, and Vishny go on to establish three propositions. First, a variety of value strategies outperformed glamour strategies over the 1968-1990 period. Second, the likely reason is that market participants appear to have consistently overestimated future growth rates of glamour stocks relative to value stocks. Third, conventional measures of fundamental risk indicate that value strategies are no riskier than glamour strategies.

Earlier it was noted that although the MV CAPM is a single-period model, theoreticians and empiricists envision a multiperiod setting where, period-by-period, investors choose MV-efficient portfolios. Unconditional tests of the CAPM are tests of the joint hypothesis that the CAPM describes asset pricing and means and betas are constant over time. Conditional tests of the CAPM allow for a dynamic environment. The papers by Jagannathan and Wang (1996) and Ferson and Harvey (1999) are excellent examples of the genre. Not only do they provide an interesting set of conditioning variables; they also provide further insight into how one might improve on the statistical properties of Fama and MacBeth's (1973) cross-sectional methodology.

Jagannathan and Wang (1996, Volume II, Chapter 9) test a conditional version of the CAPM, where betas and the market risk premium are allowed to vary over time. It is noteworthy that they include human capital in the return on aggregate wealth. But it is somewhat ironic that they assign human capital the role of a traded rather than an untraded asset, as it was in Mayers' (1972) original treatment of the subject. The article provides strong empirical evidence in support of the conditional model. However, nothing is without cost. The conditional model has three betas rather than the single-beta of the unconditional CAPM. Furthermore, Jagannathan and Wang (1996, page 9) acknowledge that the modeling exercise requires a number of 'restrictive assumptions regarding the nature of the stochastic process governing the joint temporal evolution of conditional market betas and the conditional market risk premium.' Or, as they note on page 37: 'The conditional CAPM ... is very different from what is commonly understood as the CAPM, and resembles the multi-factor model of Ross (1976).'

Ferson and Harvey's (1999, Volume II, Chapter 10) article illustrates the intimate relationship between tests of asset pricing and market efficiency, as the authors draw their conditioning variables from the return predictability literature. The paper provides a powerful rejection of the claim that Fama and French's (1993) three-factor model as well as Elton, Gruber, and Blake's (1995) four-factor model are complete models of asset pricing. Basically, Fama and French make a lot of the fact that their intercepts are zero on average. But Ferson and Harvey show convincingly that those same intercepts are significantly time varying and thus cannot be always zero. Thus, Ferson and Harvey (1999, page 1327) state: 'Our results raise a caution flag for researchers who would use the FF and Elton et al. models to control for systematic patterns in risk and expected return. Our results carry implications for risk analysis, performance measurement, cost-of-capital calculations, and other applications.'

As if the previously discussed problems that plague empirical research were not enough, there is one more. A truly disturbing aspect of empirical research is how a seemingly innocuous
change in the length of the sample period, or in the way the data are grouped, or in the precise way variables are defined, or in the exact statistical methodology employed can completely change the results. An obvious example is the "small-firm" effect. Black (1993, page 37) takes a perhaps overly strong position against what he sees as data mining in this case, arguing:

Still, it's a curious fact that just after the small-firm effect was announced, it seems to have vanished. What this sounds like is that people searched over thousands of rules until they found one that worked in the past. Then they reported it, as if past performance were indicative of future performance. As we might expect, in real-life, out-of-sample data, the rule didn't work any more.17

Unfortunately, there are many instances of these types of phenomena.

Fama and French (1992) report that there is only a weak positive relation between average return and beta over the 1941-1990 period, and virtually no relation over the shorter 1963-1990 period. Yet, Kothari, Shanken, and Sloan (1995) show that the Fama-French evidence provides little basis for rejecting the null hypothesis of a nontrivial six percent per annum risk premium over the post-1940 period.

Daniel and Titman (1997) provide evidence in favor of the hypothesis that the book-to-market characteristic is compensated irrespective of risk loadings over the 1973-1993 period. Yet, Davis, Fama, and French (2000) argue that the results are sample specific. Their evidence supports the conclusion that the three-factor risk model explains the value premium better than the characteristic model over the 1929-1997 period.

Fama and French (1996a) assert that the three-factor model outperforms the CAPM because the absolute pricing errors of the CAPM are three to five times those of the three-factor model, when portfolios are formed on the basis of firm-specific attributes. Fama and French (1995, 1996a, 1996b, 1998) and Davis, Fama, and French (2000) provide out-of-sample evidence in favor of the three-factor model. Yet, with industry data Fama and French (1997) find the absolute values of the pricing errors of the two models are about equal and Grauer (2002) finds the absolute values of the pricing errors of the CAPM are smaller.

Ferson and Harvey (1999) test a conditional form of the three-factor model. As noted above, they argue that if the slopes are time varying, it is not evidence against either the CAPM or the three-factor model; rather it is evidence that the model should be applied in its conditional form. On the other hand, if the alphas are time varying, it is strong evidence against the model. Their strong denunciation of the three-factor model is based on evidence of time-varying alphas. Yet, Grauer (2002) finds little evidence of time-varying alphas in an industry setting.

It is not clear why we get different results if we group securities by industry rather than by firm characteristics. Who knows, there might be something special about industries. But as Kothari, Shanken, and Sloan (1995, page 221) argue: 'A useful pricing model must be trusted to work under a wide variety of conditions and not just for a limited set of portfolios.'

Continuing on, Bansal, Hsieh, and Viswanathan (1993) and Ghysels (1998) suggest that the nonlinear APT is empirically more successful than the conditional CAPM as it yields smaller pricing errors. Yet, Kan and Wang (2000), using a nonparametric version of the conditional CAPM, present rather striking evidence that the conditional CAPM does a substantially better job in explaining expected returns.
Finally, an even more troubling example arises in the closely related performance evaluation literature.\textsuperscript{18} Jensen (1968) reports negative investment performance for mutual funds over the 1945-1964 period. This is taken to be strong evidence in favor of the semi-strong form of the efficient market hypothesis. Ippolito (1989) then reports positive performance for mutual funds over the 1965-1984 period. However, Elton, Gruber, Das, and Hlavka (1993) note that the performance (e.g., the Jensen's alpha) of ten passive size portfolios is negative in the Jensen period and positive in the Ippolito period. In the process of writing this introduction, I confirmed that for ten size portfolios and twelve industry portfolios, the unconditional and conditional Jensen and Fama and French performance measures (see footnote 18) are primarily negative (positive) in the 1945-64 (1965-84) period.

This is disconcerting. The profession drew a major conclusion regarding semi-strong form market efficiency. Yet, the conclusion is based on results that may be time-period (and possibly measure) specific. If time had unfolded differently so that the positive (negative) performance was recorded in the first (second) time period, would we have initially concluded that markets are semi-strong form inefficient?

It is possible, however, that I am overstating the case. An alternative interpretation is that time variation in performance evaluation may not be too surprising. Depending on the position of the market portfolio relative to the sample efficient frontier, the average mutual fund "alpha" could easily be negative sometimes and positive other times (and significantly so if the market index is within the frontier). This was suggested in Roll (1978), who showed how different index positions lead to wildly varying judgments about performance. (The examples in the next section also illustrate this point.)

V Tests of the Linear Risk Tolerance CAPMs

Rubinstein (1974) derives the linear risk tolerance (LRT) CAPMs from the LRT utility functions

\begin{equation}
  u(w_{it}) = \frac{1}{\gamma}(w_{it} + a_i)^\gamma, \quad \gamma < 1, \tag{29a}
\end{equation}

\begin{equation}
  u(w_{it}) = -(a_i - w_{it})^\gamma, \quad \gamma > 1, \text{ and } a_i \text{ large}, \tag{29b}
\end{equation}

\begin{equation}
  u(w_{it}) = -\exp(a_i w_{it}), \quad a_i < 0, \tag{29c}
\end{equation}

where $w_{it}$ is individual $i$'s wealth at time $t$, $a_i$ is the individual's risk-aversion parameter, and a common value of $\gamma$ must be shared by all investors in an economy. When all investors in an economy have the same $\gamma$, the separation property ensures they hold the same mix of risky assets, i.e., the market portfolio. An investor whose $a_i$ value is greater than the representative investor's $a$ value borrows to finance his investment in the market, while an investor whose $a_i$ value is less than the representative investor's $a$ value dampens the risk of holding the market portfolio by lending at the riskless rate. In a CAPM based on any of the polynomial families of utility functions in equation (29b), risky assets are inferior goods. Special cases include quadratic utility ($\gamma = 2$), which leads to the mean-variance CAPM, and cubic utility ($\gamma = 3$), which leads to a mean-variance-skewness CAPM. In a CAPM based on any of the power families in equation
(29a), risky assets are normal goods. In the special case where \( \gamma \) approaches 0, equation (29a) reduces to 
\[
\ln(w_i + a_i).
\]

The generalized power utility-based CAPMs are worthy of study for two reasons. First, the generalized power functions in equation (29a) provide a richer menu of investor tastes than the power functions (with \( a_i \) set equal to zero) that are typically employed in multiperiod investment and consumption theory. For example, power functions are defined on positive wealth levels. Yet, an investor may wish to protect against a loss that would leave him with wealth below some positive level. He might do so for either physiological or psychological reasons perhaps related to his place in the life cycle or to keeping up with the Jones. This type of aversion to risk may be captured with a negative \( a_i \) value in equation (29a).\(^{19}\) Second, as noted, a generalized power utility-based CAPM implies risky assets are normal goods, while a MV CAPM based on quadratic utility implies that risky assets are inferior goods.

Of course, beauty is in the eye of the beholder. Ross argues for placing restrictions on distributions rather than on preferences. For example, Ross (1978, page 888) states:

I believe there is a sound reason why explorations of utility-based CAPM theories have not attracted more effort. In any such theory, when we aggregate to the market level we must impose severe homogeneity requirements on utility functions to get results that limit data. In addition, we are also forced to impose strong homogeneity restrictions on perceived distributions; first to keep the results tractable and, second, and more importantly, to relate the theory to observables. … Given that such homogeneity is going to be imposed eventually, it would seem natural to begin the CAPM story with restrictions on distributions rather than on preferences.

Clearly, this argument, together with the APT, carried the day. But ironically, as noted, the APT may not be as preference-free as was originally thought. Moreover, preference-based theories continue to play a central role in the consumption CAPM literature.

Roll (1973) is the first to compare the performance of the MV CAPM with a 'growth optimal' CAPM, where the representative investor has logarithmic utility, i.e., \( u(w_i) = \ln(w_i) \). He finds the MV and growth optimal CAPMs are empirically indistinguishable.

Grauer (1978, Volume II, Chapter 11) pits five power utility models against the MV CAPM. He shows for the power functions in equation (29a) Rubinstein's valuation equation (6) reduces to a series of generalized SMLs

\[
E(r_j) = r_f + \left[ E(r_m) - r_f \right] \text{Risk}_j, \text{ for all } j,
\]

where \( r_j, r_m, \) and \( r_f \) are the rate of return on asset \( j \), the market portfolio, and the risk-free asset, respectively, and \( \text{Risk}_j = \text{cov}(r_j, (1 + d + r_m)' / \text{cov}(r_m, (1 + d + r_m)' ) \), with \( c = \gamma - 1 \). The parameter \( d \), which is equal to the representative investor's risk-aversion parameter \( a \) divided by initial wealth, guarantees that the representative investor neither borrows nor lends.

While Grauer (1978) is primarily an empirical paper, I believe the example in section 2.2 is more important than the empirical work as it illustrates fundamental problems associated with the traditional two-pass regression tests. I take the liberty of extending the example here: (1) by adding a fourth risky asset; (2) by adding mean – standard deviation and mean-beta / mean-Risk.
plots to the figures; and, (3) by adding Roll's (1980) orthogonal frontier, that was foreshadowed in the original example, to the mean – standard deviation plots. (Roll (1980) established that if the market is not MV efficient there are zero-beta (orthogonal) portfolios at all levels of return and these portfolios lie inside a minimum-variance orthogonal frontier.)

Table 1 contains the data needed to construct the figures for the extended example. Figure 1 depicts a generalized logarithmic utility economy with a riskfree rate of three percent. In this economy, the mean-Risk tradeoff is exactly linear. But, the mean-beta tradeoff, estimated from an OLS regression of means on betas, is negative! (Note that the ex ante expected returns, betas and Risks, and the market portfolio are observed without error.) The zero-beta portfolio estimated from the OLS regression is only one of an infinite number of zero-beta portfolios that plot inside the orthogonal frontier. Figure 2 shows that in a generalized logarithmic utility economy with a riskfree rate of nine percent, the weights in the market portfolio, the $d$ parameter, and the Risk values are quite different from those shown in Figure 1. Moreover, the mean-beta tradeoff estimated from the OLS regression is now positive and the new OLS zero-beta portfolio plots inside a different orthogonal frontier. Figure 3 depicts a mean-variance economy with a riskfree rate of three percent. In this economy, the tangency portfolio is the market portfolio, the mean-beta tradeoff is exactly linear, the mean-Risk tradeoff (estimated from an OLS regression of means on Risks) is not, and the OLS zero-beta portfolio's expected return matches that of the riskfree asset. Figure 4 portrays a mean-variance economy with a riskfree rate of nine percent. Naturally, the new mean-beta tradeoff is exactly linear, and the new mean-Risk tradeoff (estimated from an OLS regression of means on Risks) is not. Figure 5 shows that neither model fits when an equal-weighted portfolio is employed as a proxy for the market portfolio.

Table 1 and Figures 1-5 here

The example highlights five sets of problems associated with the traditional two-pass regression tests. The first set of problems is associated with two well-known facts. If the market portfolio is MV efficient, then: (1) securities plot on the SML, and (2) the rates of return on all zero-beta portfolios are equal to the riskfree rate. Figures 1 and 2 show that, if the market portfolio is not mean-variance efficient, then securities do not plot on the SML and there are numerous zero-beta portfolios whose expected rates of return are not equal to the riskfree rate. The results are interesting for two reasons. First, Black (1972) developed the zero-beta version of the MV CAPM having observed Black, Jensen, and Scholes' (1972) result that the zero-beta rate estimated over long periods of time exceeds the average riskfree rate. Had he paid more attention to Black, Jensen, and Scholes' sub-period results, where in some cases the estimated zero-beta rate is less than the average riskfree rate, he may have correctly concluded that the empirical results are more consistent with an inefficient market portfolio. Second, cross-sectional tests of the MV CAPM judge whether the model holds by comparing the zero-beta rate from an OLS regression of means on betas with the riskfree rate of interest. But these tests are ambiguous because the zero-beta rate estimated from the OLS regression is only one of an infinite number of zero-beta rates if the market is not MV-efficient. (Note that, in general, the OLS estimate of the zero-beta portfolio does not even lie on the minimum-variance orthogonal frontier.)

The second problem is that in the generalized power utility economies the expected return-Risk plot is not linear unless we correctly identify the market portfolio, $\gamma$, and $d$. For example, when unity plus the riskless rate of return is 1.09 in the generalized logarithmic utility economy,
\( d = 0.269 \), and \( \text{Risk}_j = \text{cov}\left( r_j, (0.731 + r_m)^{-1}\right) / \text{cov}\left( r_m, (0.731 + r_m)^{-1}\right) \). A regression of means on Risk confirms the linear mean-Risk plot with an intercept of 1.09, a slope of 0.19221, and an \( R^2 \) equal to 1. If we mistakenly set \( d \) equal to 0, which is consistent with the representative investor holding the 'growth-optimal' portfolio, then \( \text{Risk}_j = \text{cov}\left( r_j, (1 + r_m)^{-1}\right) / \text{cov}\left( r_m, (1 + r_m)^{-1}\right) \). In this case a regression of means on Risks has an intercept of 1.052, a slope of 0.2312, and an \( R^2 \) of 0.94. In a sense, these results foreshadow problems with the consumption-based CAPM, where the early empirical work assumes power utility with the \( d \) parameter set equal to zero. This raises two possibilities. The power utility framework, with \( d \) set equal to zero, may not be rich enough to describe real-world tastes. Alternatively, the returns in the example take on a wide range of values. Consumption data may not vary enough to allow us to distinguish between different models.

Third, the example illustrates the fundamental importance of identifying the exact composition of the market portfolio. Or stated another way, the example serves as a cautionary tale for a profession whose basic empirical methodology consists of judging a model by testing whether average returns are equal to the expected returns predicted by the model. In a generalized logarithmic economy, Risk is the correct (and beta is the wrong) measure of risk. In a MV economy, beta is the correct (and Risk is the wrong) measure of risk. But given a riskfree interest rate, the relative prices (the weights in the market portfolio) in the two economies are quite different. Basically, there is a one-to-one correspondence between a mean-Risk or a mean-beta tradeoff (stochastic discount factor) and the relative prices (the weights in the market portfolio) in the two economies. Naturally then, figure 5 shows that with an equal-weighted portfolio as the proxy for the market portfolio, neither model fits the data. In light of this, it makes little or no sense to argue—as I have and we do as a profession— that the model exhibiting the smallest pricing errors is the best model, unless we employ the true market portfolio in the tests.

Fourth, almost all empirical work employs excess returns assuming that this gets around the problem of a changing riskfree rate of interest. The changes that occur in the composition of the market portfolio, the market portfolio's mean and variance, and the betas (Risks) when the interest rate changes in a mean-variance (generalized logarithmic utility) economy indicate that this is a heroic assumption. (See the plots of the mean-variance economies in Figures 3 and 4 and the generalized logarithmic utility economies in Figures 1 and 2.)

Fifth, there are many stochastic discount factors consistent with different sets of investor tastes in equation (29a) and different inefficient portfolios being the market portfolio. In these LRT economies, the representative investor neither borrows nor lends. However, an investor with power utility and with \( a_i = 0 \) may take large positive or negative positions in the riskfree asset. For example, in results not reported in the figures, there is an equilibrium expected return - Risk tradeoff for an economy with a nine percent riskfree rate, \( \gamma = -50 \), and \( d = 38.11 \). In this economy, a \(-50\) power investor (with \( a_i = 0 \)) lends over 97 percent of his wealth. On the other hand, in the generalized logarithmic utility economy with a nine percent riskfree rate, an investor with logarithmic utility (and \( a_i = 0 \)) borrows approximately 0.33 of his wealth (and when the interest rate is three percent in the generalized logarithmic utility economy, this investor borrows 2.75 times his wealth). Again the moral is obvious. The family of power functions, with \( a_i = 0 \), may not be rich enough to describe asset pricing.
As an aside, it is worth noticing the similarity of some aspects of the example to the results found in two sets of papers by Roll (1978) and Grauer (1991), which discuss performance evaluation, and by Roll and Ross (1994), Kandel and Stambaugh (1995), and Grauer (1999), which discuss possible shortcomings in the tests of the CAPM. Both sets of papers highlight the dangers of focusing exclusively on mean-beta space. From the point of view of performance measurement, the message is that a market index can be quite close to the efficient frontier. Yet, different index positions can lead to wildly varying judgments about performance. From the point of view of asset pricing, the market index can be almost efficient, yet the slope of the SML can be zero. Or, the market index can be grossly inefficient, yet the mean – beta plot can be almost exactly linear.

While I strongly believe in the value of examples in general and this one in particular, it should be noted that others don't. The referee of the original 1978 paper had two reactions to the example. First, he wrote in the margin: 'This guy is as mixed up as Roll', which—contrary to his meaning—I take as high praise indeed. Second, he recommended that the example be deleted. This is all quite unremarkable—except that Eugene Fama was the referee. We may never know whether Fama values examples any more now than then. On the other hand, the following statements from Fama and French (1996b, page 1956) suggest that observing the true market portfolio, for example, has not risen to the top of his research agenda:

It is, of course, possible that the apparent empirical failures of the CAPM are due to bad proxies for the market portfolio. In other words, the true market is mean-variance-efficient, but the proxies used in empirical tests are not. In this view, revival of the CAPM awaits the coming of $M$. …This bad-market-proxy argument, however, does not justify the way the CAPM is currently applied … If the common market proxies are inefficient, then applications that use them rely on the same flawed estimates of expected return that undermine empirical tests of the CAPM. Like the redemptive empirical tests, valid applications of the CAPM await the coming of $M$.

The major empirical result in Grauer (1978) is that the data do not allow us to distinguish between the models. It is hard to say why this occurs. It may be that, with monthly data, returns are close enough to being normally distributed or sufficiently 'compact' to ensure that all risk-averse investors are 'almost' MV decision-makers. Or, it may be that the models really are different and the cross-sectional regression methodology, accentuated by the choice of an equal-weighted market proxy, is the source of the problem.

One might hope that portfolio theory could shed light on whether the models differ. Unfortunately, the results are ambiguous. Grauer (1981) repeats the generalized SML tests over a broader range of powers ranging from $\gamma = 0.5$ to -30 using both a value-weighted and equal-weighted proxy for the market. Again, the data do not allow us to distinguish between the models. But, the investment policies of the power models are different from each other and from the MV model. On the other hand, Grauer and Hakansson (1993) find that, with quarterly decision horizons, the out-of-sample returns earned from portfolios based on MV approximations to power utility functions, with $T$ set equal to $1/(1-\gamma)$ in equation (1), are fairly close to the returns earned by the power utility portfolios themselves.

VI Tests of the Consumption-Based CAPM

As noted in section I, the consumption-based CAPM makes both time-series and cross-sectional predictions. Both predictions may be tested following the seminal works of Hansen
and Singleton (1982, 1983), where estimation is based on Hansen's (1982) generalized method of moments. For the most part, the tests reject the CCAPM. The inability of the model to match seemingly reasonable levels of risk aversion with the observed volatility of consumption growth (discussed in section I) is particularly troubling. Even so, this is only one of three major puzzles that the time-separable power utility CCAPM cannot explain: Mehra and Prescott's (1985) equity premium puzzle, Weil's (1989) riskfree rate puzzle, and Backus, Gregory, and Zin's (1989) term premium puzzle. On a much more positive note, attempts to explain the 'puzzles' have generated a richer set of models. For example, the habit formation models of Abel (1990), Constantinides (1990), and Campbell and Cochrane (1999) attempt to explain the equity premium by formulating a model in which utility depends on past consumption.

The central cross-sectional prediction of the CCAPM is that expected returns are linearly related to consumption betas. It is encouraging that Breeden, Gibbons, and Litzenberger (1989, Volume II, Chapter 12) cannot reject this hypothesis with U.S. data. Wheatley (1988a) reports confirmatory results with U.S. data and Wheatley (1988b) cannot reject the linearity hypothesis with international data.

Breeden, Gibbons, and Litzenberger also discuss some of the econometric difficulties associated with consumption data. Ferson and Harvey (1992) take this discussion a step further focusing on the smoothness in the growth of consumption expenditures relative to stock market returns. Part of the apparent smoothness comes from the way consumption data are reported. Ferson and Harvey document a startling difference in the variability of seasonally adjusted consumption data (analyzed in most papers) and raw consumption data. Somewhat surprisingly, the model does not fit the raw consumption data all that well either.

Campbell (2000, Volume II, Chapter 13) surveys the field of asset pricing with an emphasis on the aspects of the field that I have not emphasized. Specifically, he focuses on the CCAPM, the stochastic discount factor, and the interplay between the stylized facts (empirical puzzles) documented in the CCAPM literature and the theory, which has been developed to explain them. As in the case for the MV CAPM - anomalies literature discussed above, this interplay between theory and empirical work adds both richness and a sense of excitement to the literature. In addition, Campbell's survey contains sections on prices, returns and cash flows and behavioral finance that I have not discussed in this review.

Finally, lest the reader think that I am giving short shrift to the CCAPM, it should also be noted that Roll and Ross (1986, Volume II, Chapter 14) and Cochrane (1996, Volume II, Chapter 17) contain tests that pit the CCAPM against the MV CAPM, the APT and an investment-based CAPM. In addition, Fama (1991) provides his views about the relative performance of tests of the CCAPM as part of his review of efficient markets and tests of asset pricing models.

VII Tests of the Arbitrage Pricing Theory

Testing the APT is tricky. Arbitrage arguments can only be used to provide an approximate factor pricing equation for some unknown number of unidentified factors. Shanken (1982), however, argues that testing requires an exact pricing equation, which in turn requires additional assumptions. As noted in section I, one way to produce an exact factor pricing equation is to formulate a competitive equilibrium model, with all the attendant assumptions about investors' tastes added back into the model. Moreover, Shanken contends that the role of the (unobservable) market portfolio in Conner's (1984) exact pricing equation may preclude
meaningful tests of the APT in the same way Roll (1977) maintains that the inability to observe the true market portfolio preclude tests of the MV CAPM.

The exact factor pricing model issue is further complicated by the fact that Merton's (1973) ICAPM, with its assumptions concerning the stochastic process, also yields an exact multifactor pricing equation. See, for example, Fama (1996) who links Merton's ICAPM, exact factor pricing, and constrained MV decision-making. In this case, the market portfolio serves as one factor and investors' demands to hedge against shifts in investment opportunities account for the other factors. Furthermore, Campbell, Lo, and MacKinlay (1997)'s discussion of multifactor models, which includes the APT, the ICAPM, and the Fama-French three-factor model, is framed exclusively in terms of exact factor pricing. But, what are we testing with exact factor pricing models: the MV CAPM, the APT, or the ICAPM? Many papers, including the series of Fama and French papers discussed above, explicitly recognize that it is not clear whether they are testing the MV CAPM (by adding variables other than beta to the SML) or some unspecified version of the APT or ICAPM.

Like many of issues in empirical finance, the contention that testing the APT requires an exact pricing equation is open to debate. First, Dybvig (1985) and Grinblatt and Titman (1983) argue that, given a reasonable specification of the parameters of the economy, theoretical deviations from exact factor pricing are likely to be negligible. Hence, they conclude that we may not need to rely on equilibrium-based derivations of the APT. Dybvig and Ross (1983) and Shanken (1985) debate the issue. Second, Richard Roll in discussing these issues with me, put forth a series of arguments to support the contention that the APT could be rejected without having to rely on exact factor pricing. This explains why, from the very earliest APT tests, it was always considered essential to check the cross-sectional relation between sample mean returns and factor betas by including additional regressors that should not be there (according to the APT). The prime candidate which should not be there is a beta against a market index. But own variance and other variables were also tried. Another approach was to test the equality of intercepts across groups. (Roll and Ross (1980) provide examples of these types of tests.) Of course, one can never prove the absence of arbitrage, but one could certainly demonstrate its existence and hence reject the APT. Imagine, for example, that one could construct a hedge portfolio (long one portfolio and short another), which had positive performance every period. That would reject the APT.

Leaving these arguments aside, there are two basic approaches to testing the APT: statistical and theoretical. Statistical approaches use factor analysis to extract the common factors and then test whether the expected returns are explained by the cross-sectional loadings of security returns on the factors. The theoretical approaches specify variables that are correlated with asset returns and test whether the loadings of returns on these economic factors explain the cross-section of expected returns. In turn, the theoretical approaches are of three main types. The first, initiated by Chen, Roll and Ross (1986), specifies macroeconomic and financial market variables that are thought to capture the systematic risks of the economy. A second method, characterized by the series of Fama and French papers discussed above, constructs hedge portfolios with long/short positions in firms with attributes known to be associated with mean returns. For instance, the SML or HML variables discussed earlier reflect the differences in returns on hedge portfolios associated with firm size and book-to-market equity ratios, respectively. A third method is often associated with the risk management firm Barra. Rosenberg (1974) and Rosenberg and Guy (1976) contain early descriptions of the technique. Basically, the method identifies firm
descriptors that may help in predicting changing measures of risk. The firm descriptors are quite extensive. Examples include: (1) descriptors of market variability (historical beta, historical standard deviation, and share turnover); (2) descriptors of earnings variability; (3) descriptors of unsuccess and low valuation (relative strength and book-to-market equity); (4) descriptors of immaturity and smallness (total assets and market value of equity); (5) descriptors of growth orientation (payout ratios, dividend yields, normalized earnings-price ratios); (6) descriptors of financial risk (debt-to-asset ratios, book leverage, and market leverage). With the descriptors identified, a time series of cross-sectional regressions are run whose coefficients presumably mimic the underlying risk factors.

Roll and Ross (1980, Volume II, Chapter 14) pioneered the use of factor analysis in testing the APT. They report the theory is supported in that at least three and probably four "price" factors are found in the generating process of returns. As noted above, they also check the cross-sectional relation between sample mean returns and factor betas by including additional regressors such as own variance that should not be there according to the APT. But, unfortunately, we do not have the data to provide an estimate of the full covariance matrix of returns and must work with groups of securities, which leads to an irresolvable debate as to the number of common factors. See the interchange between Dhrymes, Friend, and Gultikin (1984), Roll and Ross (1984), and Dhrymes, Friend, and Gultikin, and Gultikin (1985), for example.

Lehmann and Modest (1988) and Connor and Korajczyk (1988) are often cited as being among the best papers that rely on statistical approaches. Lehmann and Modest (1988) employ more efficient computational maximum-likelihood factor analytic methods that enable them to study larger data sets. Their results show little sensitivity when the number of factors increases from five to ten to fifteen. However, they are unable to explain the returns of portfolios formed on the basis of firm size (the market value of equity). Connor and Korajczyk (1988) propose an asymptotic principal components approach. They propose using the eigenvectors associated with the $k$ largest eigenvalues of a $(T \times T)$ centered returns cross-product matrix rather than the standard approach, which uses the principal components of the $(N \times N)$ sample covariance matrix. The advantage of this approach is that they are able to work with a far smaller matrix, as $T$—the number of time series observations—is typically far smaller than $N$—the number of assets. Their empirical techniques allow for fairly arbitrary time variation in risk premiums. They report that the APT provides a better description of the expected returns on assets than the MV CAPM. However, the APT still allows for some statistically reliable mispricing of assets.

Chen, Roll, and Ross (1986, Volume II, Chapter 15) test whether macroeconomic variables are risks that are rewarded in the stock market. I find this approach far more interesting than the statistical approach. Like Fama (1991), I am more interested in what the factors are, and what their economic meaning is, rather than in how many factors there are. The variables Chen, Roll, and Ross include are the growth rate of industrial production, the difference between the returns on high and low-grade bonds (a risk premium), the difference between the returns on long and short-term bonds (a term premium), and unexpected inflation. They find these sources of risk (especially the first three) are significantly priced. Moreover, it is striking that compared to the economic state variables, betas calculated from value-weighted and equal-weighted NYSE indices and consumption betas have little influence on pricing. Nor does an index of oil prices.

It is worth noting that Chen, Roll and Ross employ a version of Fama and MacBeth's (1973) cross-sectional approach to testing. This approach focuses on whether the factors are priced, i.e., on whether the gammas in equations (20) or (21) are statistically different from zero. Fama and

Shanken (1982, Volume II, Chapter 16) challenges the view that the APT is more susceptible to empirical verification than the CAPM. Specifically, he argues against the use of factor analytic techniques. First he shows that equivalent sets of securities (that is, securities that yield identical portfolio returns) may conform to very different factor structures. Then he shows that the usual empirical formulation of the APT, when applied to these structures, may yield different and inconsistent implications concerning expected returns for a given set of securities. The implications will be consistent if and only if all of the securities have the same expected return. Ironically, the usual formulation of the theory precludes the expected return differentials it attempts to explain.

VIII Tests of an Investment-Based CAPM Using the Generalized Method of Moments

Cochrane (1996, Volume II, Chapter 17) examines an investment-based asset-pricing model. He finds that the investment-based model performs about as well as the CAPM and the Chen, Roll, and Ross (1986) model, and substantially better than a simple consumption-based asset pricing model. While the results are of interest in and of themselves, the discussion of the methodology is absorbing. Cochrane casts the models in a stochastic discount factor framework: \( p = E(mr) \), where \( p \) is a vector of prices, \( E(.) \) is the expectations operator, \( m \) is the stochastic discount factor, and \( r \) is a vector of returns. (The "price" of a return is one.) He then provides a very readable description of a generalized method of moments (GMM) test of factor pricing models. All in all, the paper makes a compelling case for the stochastic discount factor - GMM methodology.

Kan and Zhou (1999, Volume II, Chapter 18) compare the stochastic discount factor method to more traditional methods of testing asset-pricing models. If a linear factor model generates asset returns, then ignoring the full dynamics of asset returns causes two potential problems. The first problem is the accuracy of the parameter estimates can be poor. In a surprising result, they show the standard error of the estimated risk premium is often more than 40 times greater than that of the traditional methodologies. The second problem with the stochastic discount factor methodology is that its specification test has very low power against misspecified models, especially when the proposed factors are not highly correlated with the returns.

Needless to say the results attracted a great deal of attention. And perhaps not surprisingly Cochrane (2001a, b) and Jagannathan and Wang (2001) have taken issue with them. The basic argument is that the Kan-Zhou results arise from a "strange" assumption. Kan and Zhou allow the traditional maximum likelihood procedure to know the factor risk premium, while the generalized method of moments – stochastic discount factor procedure must estimate it. Cochrane and Jagannathan and Wang then show how this assumption leads to the dramatic increases in the standard errors of the generalized method of moments technique. I find the exchange most interesting. It is not hard to imagine how the assumptions needed to apply empirical techniques may not be relevant in the real world and that this could lead to erroneous inferences. It is much harder to imagine that the results of a controlled Monte Carlo experiment could be so different. Perhaps this is a cautionary tale for us all.
IX Summary

Although they are sometimes technically demanding, classical theories of asset pricing are really rather simple. For the most part, they are demand driven. The pricing equations are the first-order conditions of a representative investor at a point in time. Sometimes the models are cast in a single period. Other times they are linked with a continuous-time stochastic process describing how returns evolve through time. As we move forward, we might hope that it would be possible to build models that link these demand conditions with supply conditions generated from the real economy. The addition of supply conditions would hopefully be coupled with a more complete consideration of transactions costs, liquidity, and taxes. In this more complete model, the stochastic process describing security returns would not be assumed. Rather, the stochastic process would be determined by supply and demand, and the time at which investors choose to rebalance their portfolios would be stochastic. The fundamental questions are: Can we build such a model? And if we can, will it describe security pricing? Or, as has been the case in the anomalies literature, must we tailor our theories more in the direction of investor irrationality and behavioral finance?

On the empirical side, I have emphasized the controversy surrounding the heated, intensely interesting debates surrounding how to structure the tests and how to interpret the results. (This volume is after all part of a series entitled: 'The international library of critical writings in financial economics'!) There are as many reasons to be optimistic as pessimistic about empirical research. Thus, you may agree with Roll's (1977, page 129) condemnation of tests of the mean-variance CAPM: '(a) No correct and unambiguous test of the theory has appeared in the literature, and (b) there is practically no possibility that such a test can be accomplished in the future.' On the other hand, you may believe that Fama and French's (1992) results mean the CAPM is dead, only to be replaced with Fama and French's (1993) three-factor model. Or, you might believe that Fama's (1976, page 370) statement: 'In truth, all we can really say at this time is that the literature has not yet produced a meaningful test of the Sharpe-Lintner hypothesis' rings just as true today as it did a quarter of a century ago.

It is easy to be pessimistic. We have never really successfully addressed the most basic issues in designing tests of our models. In the case of the mean-variance and related models we chose to test risk-return tradeoffs using statistical techniques in spite of the fact that casual observation indicates other implications of the model—two-fund separation, for example—are not true. Furthermore, we ignore numerical examples, which show that even slight perturbations from a true risk-return relationship cause other implications of a model—such as the existence of positive prices—to be false.

If we are to test risk-return tradeoffs, we must establish a link between ex ante beliefs and ex post observations. The most obvious way to do so is to assume that investors have homogeneous beliefs. Yet, that would seem to preclude trade. We assume expectations are realized, at least on the whole and on the average. But, as Elton (1999) notes, even casual observations suggest that they may not be. The early tests of the mean-variance CAPM assume returns and betas are stationary. This implies that the riskfree rate never changes and unbelievably that prices never change! More troubling if possible, Cheng and Grauer (1982) suggest that the models themselves may rule out the possibility that ex post returns are drawn from ex ante return distributions.
Roll (1977) makes the compelling argument that the mean-variance theory (and by extension many of its competitors) is not testable unless the market portfolio is observable and used in the tests. Still, we rush to test using proxies for the market. Roll and Ross (1994), Kandel and Stambaugh (1995), and Grauer (1978, 1999) present relatively simple examples that show how the tests might fail. But, we continue to test.

Aeronautical engineers determine whether model aircraft fly in wind tunnels before they attempt to fly the real thing. Physicists test the laws of gravity in a vacuum, not in the midst of a hurricane. Maybe we can learn from them. No one has explicitly said that financial economists should demonstrate that a test design works using simple numerical examples before they test a model with real-world data. But maybe someone should.

Empirical work emphasizes hypothesis testing. But, maybe we should pay more attention to the logic of the test design than to the statistical properties of a test. McCloskey and Ziliac (1996, page 112) state:

No economist has achieved scientific success as a result of a statistically significant coefficient. Massed observations, clever common sense, elegant theorems, new policies, sagacious economic reasoning, historical perspective, relevant accounting: these have all led to scientific success. Statistical significance has not.

Given the obvious flaws in the test designs employed in the asset pricing literature, it is easy to see why.

Finally, data snooping calls into question much of the empirical work surrounding the mean-variance CAPM in the 1990's. But, we should be careful. I believe that it is an oversimplification to think that theory should come before testing.22 Leamer (1996, page 189) is more blunt: 'As you wander through the thicket of models, you may come to question the meaning of the Econometric Scripture that presumes the model is given to you at birth by a wise and beneficent Holy Spirit.' Rather, the process is much more subtle. There is an undeniable interaction between theory and empirical work that leaves room for a much more optimistic view of empirical work. If for the moment we cast aside doubts about test design, it is not hard to conclude that empirical work concerned with market efficiency and asset pricing, with its attendant anomalies and empirical puzzles, has led to a deeper understanding of the cross-sectional and time-series characteristics of security returns.

Of course, it will be interesting to learn which of the anomalies and puzzles will stand the test of time and which will turn out to have been time period specific. Moreover, it will be interesting to see whether we can build rational theories to explain this richer empirical environment or whether we will turn to behavioral theories to lead us out of the morass.

To sum up, the lack of any clear-cut understanding of which theories may prevail is discouraging; but the richness of the debates surrounding the empirical tests of the theories together with the interaction of theory and empirical work is truly invigorating. To put things in perspective, we need only recall the old adage that the journey is more important than the destination. For modern theories of asset pricing, the journey began less than forty years ago. Perhaps we are closer to our destination of understanding how assets are priced, perhaps not. Perhaps we have traveled the right roads, perhaps not. To me, however, it doesn't really matter how far I believe we have come, or how far we have to go, what matters is that, so far at least, it has been one hell of a ride.
Footnotes

1 A number of textbooks provide excellent treatments of these issues. See Ingersoll (1987) or Cochrane (2001a), for example.

2 Campbell, Lo, and MacKinlay (1997) and Cochrane (2001a) discuss the different empirical methods in detail.

3 Although it is often claimed that the assumption of jointly normal return distributions implies that any expected utility maximizing agent will pick an MV efficient portfolio, the utility functions must be defined over all wealth levels. Therefore, power functions—arguably the most interesting utility functions—are excluded as they are only defined on positive wealth.

4 One may view the objective function in equation (1) in one of two ways. First, one could simply assert that the investor trades off mean against variance. Or, one could argue that the objective function is a Taylor-series approximation to an investor’s true expected utility problem.

5 If there is no riskless asset, \( R_f \) is replaced by \( R_z = \lambda / T_m \), the return on a zero-beta portfolio.

6 A more complete model would also be concerned with the productive possibilities in the economy. For example, Cochrane (2001a) and Cochrane (1996), discussed below, show the stochastic discount factor framework is general enough to encompass asset pricing based on portfolio choice models, factor models, or production models. In addition, a more complete model would explain how the interplay of supply and demand determines the stochastic process that describes how prices evolve rather than simply assuming it as the current generation of multiperiod models does.

7 The numbers change from data set to data set. For example, Campbell, Lo, and MacKinlay (1997) who employ a longer time series of data present estimates of the same variables that are a little less extreme.

8 However, there is a question as to whether generalized least squares regressions should be run at each point in time. See Huang and Litzenberger (1988) and Cochrane (2001a) for detailed discussion of the approach. In addition, Jagannathan and Wang (1996), Kan and Zhang (1999a), and Ferson and Harvey (1999) provide further discussion of this point.

9 Fama’s dynamic programming approach is the most general treatment of the discrete-time portfolio selection problem. However, at the risk of beating a dead horse, I note that the last step, where he introduces normality to justify the MV model, holds only if consumers are willing to tolerate the possibility of negative consumption.

10 Ross (1978) is well aware of these and related problems. On page 897 he states: ‘There is a strong sense in which individual behavior in the lognormal case, for example, is inherently irrational. At each moment of time the market portfolio is efficient, hence with no stochastic changes in the underlying description of anticipated returns, the same market portfolio (or a deterministically different one) will be chosen at all points of time. But, since the proportions in the market portfolio are total values, with exogenously given supplies, asset prices will also remain unchanged over time (or change deterministically). This means that only dividend uncertainty can exist in equilibrium; capital gains are sure. Or, in the absence of dividend payouts, as in the simpler models, all returns
are ex ante viewed as random (lognormally distributed), but ex post prices grow at a certain (identical) rate and returns must be sure.'

11 More recently, Moskowitz and Grinblatt (1999) document a strong and prevalent momentum effect in industry components of stock returns, which accounts for much of the individual stock momentum anomaly.

12 It is somewhat surprising that this particular result generates so much attention, in light of the fact that years earlier Black, Jensen, and Scholes (1972) report at least half a dozen instances of negative risk-return tradeoffs estimated over two and eight year periods.

13 Note, however, if one runs a regression of portfolio returns on portfolio betas, where the portfolios are sorted by size alone (with the data are taken from the first two lines of Fama and French's Table II), the slope of the SML is too steep. For further discussion of problems associated with grouping, see the summary of Lo and MacKinlay (1990) and footnote 15 below.

14 Arguably, useless factors might be more closely associated with macroeconomic variables than with portfolios generated by grouping on firm characteristics.

15 It is disturbing that the outcome of the tests is affected by grouping. Roll (1977) suggests that grouping might cause the CAPM to appear to be true when it is false. Berk (2000) analyzes the theoretical implications of sorting data into groups and then running the tests within each group as, for example, Daniel and Titman (1997) do. He shows that, by picking enough groups to sort into, a researcher can destroy the within-group explanatory power of a correct asset-pricing model. Grauer and Janmaat (2002) identify a number of unintended consequences of grouping when the capital asset pricing model is true and when it is false. When the model is true, traditional grouping can cause problems with the most basic capital asset pricing and cross-sectional regression relationships. For example, with traditional grouping, the market portfolio is super-efficient—unless securities in each group are value-weighted. Yet, when the model is grossly false, grouping can cause the model to appear to be absolutely correct. Ironically, the only way this can occur is when securities in each group are value-weighted. To make matters worse, when the model is false, the slope of a cross-sectional regression of expected returns on betas fitted to grouped data may be flatter than when the regression is fitted to ungrouped data, thereby exacerbating the very problem grouping was meant to alleviate.

16 Nonetheless, they are still tests of a joint hypothesis that replace the assumption of return stationarity with specific often times interesting and innovative but ad hoc assumptions about how means, variances, covariances, market risk premiums, betas, or even alphas evolve through time. Moreover, for the most part, the tests do not address the problem of using a proxy for the market portfolio or the problem that in a CAPM world realized returns are not drawn from the ex ante distributions envisioned by investors.

17 A more charitable explanation is that the small-firm premium was bid away as investors became more aware of it.

18 As Grauer (2002) notes, performance measures are so intimately related to models of asset pricing that one could argue time-series tests of asset pricing models are just a way of benchmarking

19 Although the modeling is somewhat different, these ideas foreshadow the habit-formation models that attempt to explain the equity premium puzzle, which bedevil the CCAPM based on power utility functions.

20 Recall the quotes from Cochrane (2001a) and Fama and French (1996a), for example.

21 While we may disagree on the value of the example, I hasten to add that few of my papers have been so competently refereed.

22 Kennedy (2001) contains an excellent discussion of the unpleasant realities applied econometricians face in working with real-world data.
References

Abel, Andrew B. (1990), 'Asset prices under habit formation and catching up with the Jones', *American Economic Review*, 80, 38-42.

Amihud, Yakov (2000), 'Iliquidity and stock returns: Cross-section and time-series effects', manuscript New York University.


Best, Michael J. and Robert R. Grauer (1990), 'The efficient set mathematics when mean-variance problems are subject to general linear constraints', "*Journal of Economics and Business*, 42, 105-120.


Cochrane, John H. (2001b), 'A rehabilitation of stochastic discount factor methodology', manuscript University of Chicago.


Kan, Raymond and Kevin Q. Wang (2000), 'Does the nonlinear APT outperform the conditional CAPM?', manuscript University of Toronto.


Klein, Peter E. (2001), 'Sinning in the basement: What are the rules? The ten commandments of applied econometrics', manuscript Simon Fraser University.


Mossin, Jan (1966), 'Equilibrium in a capital asset market', *Econometrica*, 34, 768-783.


Rosenberg, Barr and James Guy (1976), 'Prediction of beta from investment fundamentals,' *Financial Analysts Journal*, 32 (May/June), 60-72 and 32 (July/August), 62-70.


Each investor in an economy sees the *ex ante* return distribution in panel A. For generalized power utility economies we solve for the weights in the market portfolio using a nonlinear programming algorithm. In a mean-variance economy we can solve for the weights in the market portfolio using the *ex ante* return distribution in panel A or the mean vector and covariance matrix in panel B. Once we have the weights in the market portfolio we can solve for the remainder of the parameters needed to draw figures 1-5.

### Panel A: The *ex ante* return distribution

<table>
<thead>
<tr>
<th>Security $j$</th>
<th>State 1</th>
<th>State 2</th>
<th>State 3</th>
<th>State 4</th>
<th>State 5</th>
<th>State 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
<td>1.</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>0.6</td>
<td>1.8</td>
<td>0.6</td>
<td>1.8</td>
<td>0.6</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>1.68</td>
<td>1.68</td>
<td>1.68</td>
</tr>
<tr>
<td>4</td>
<td>0.</td>
<td>0.</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

| Probability  | 0.05    | 0.05    | 0.4     | 0.4     | 0.05    | 0.05    |

### Panel B: The mean vector and covariance matrix

<table>
<thead>
<tr>
<th>Security $j$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Vector</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.150</td>
<td>1.200</td>
<td>1.255</td>
<td>1.350</td>
<td></td>
</tr>
<tr>
<td><strong>Covariance Matrix</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.202500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.</td>
<td>0.360000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.063750</td>
<td>0.204000</td>
<td>0.180625</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.022500</td>
<td>0.</td>
<td>0.063750</td>
<td>0.202500</td>
</tr>
</tbody>
</table>
Figure 1. A generalized logarithmic utility economy with a riskfree rate of three percent. In this economy, the weights in the market portfolio are (0.58, 0.02, 0.28, 0.11) and the value of \( d \) that guarantees the representative investor neither borrows nor lends is equal to -0.756. The mean-Risk tradeoff is exactly linear. But the mean-beta tradeoff, estimated from an OLS regression of means on betas, is negative. The zero beta portfolio estimated from the OLS regression is only one of an infinite number of zero-beta returns that plot inside the orthogonal frontier.

**Symbols**: \( T \) is the tangency portfolio. \( E \) is the equal-weighted portfolio. \( M \) is the market portfolio. \( ZB \) is the zero-beta portfolio estimated from an ordinary least squares regression of means on betas. \( Rf \) is unity plus the riskless rate of interest. The orthogonal frontier contains minimum-variance portfolios whose returns are orthogonal to (or equivalently have zero beta relative to) the market portfolio's returns. SML is the equilibrium pricing relation between means and betas in a mean-variance economy (or means and Risks in a generalized logarithmic utility economy). OLS Beta (OLS Risk) is the line representing an OLS regression of means on betas (means on Risks). The assets are labeled 1-4 in mean-standard deviation space, B1-B4 in mean-beta space, and R1-R4 mean-Risk space.
Figure 2. A generalized logarithmic utility economy with a riskfree rate of nine percent. In this economy, the weights in the market portfolio are (0.03, 0.01, 0.63, 0.33) and the $d$-parameter is equal to -0.269. The mean-Risk tradeoff is exactly linear. But now the mean-beta tradeoff, estimated from an OLS regression of means on betas, is positive. The OLS zero-beta portfolio plots inside a different orthogonal frontier than the one shown in Figure 1. For the definition of the symbols see Figure 1.
Figure 3. A mean-variance economy with a riskfree rate of three percent. In this economy, the tangency portfolio is the market portfolio with weights equal to (0.16, 0.16, 0.06, 0.62). The risk measures employ a value of \( d \) equal to -0.756. The mean-beta tradeoff is exactly linear. The mean-Risk tradeoff, estimated from an OLS regression of means on Risks, is not. For the definition of the symbols see Figure 1.
Figure 4. A mean-variance economy with a riskfree rate of nine percent. In this economy, the tangency portfolio is the market portfolio with weights equal to (0.02, 0.04, 0.25, 0.69). The risk measures employ a value of $d$ equal to -0.269. Again, the mean-beta tradeoff is exactly linear whereas the mean-Risk tradeoff, estimated from an OLS regression of means on Risks, is not. For the definition of the symbols see Figure 1.
Figure 5. The problem of using an equal-weighted portfolio as a proxy for the market portfolio. In this example, the mean-Risk and mean-beta points are calculated assuming that the equal-weighted is the portfolio market portfolio. Neither model fits in this case. For the definition of the symbols see Figure 1.